80/262/CDV

COMMITTEE DRAFT FOR VOTE (CDV)
PROJET DE COMITÉ POUR VOTE (CDV)

	Project number Numéro de projet	80/61162-410/Ed. 1
IEC/TC or SC: **80** CEI/CE ou SC:	Date of circulation Date de diffusion **2000-04-07**	Closing date for voting (Voting mandatory for P-members) Date de clôture du vote (Vote obligatoire pour les membres (P)) **2000-09-15**

Titre du CE/SC:	TC/SC Title: Maritime navigation and radiocommunication equipment and systems
Secretary: M. A. RAMBAUT - United Kingdom Secrétaire:	
Also of interest to the following committees Intéresse également les comités suivants	Supersedes document Remplace le document 80/175/CD & 80/176/CD

Horizontal functions concerned
Fonctions horizontales concernées

□ Safety Sécurité	□ EMC CEM	□ Environment Environnement	□ Quality assurance Assurance qualité

CE DOCUMENT EST TOUJOURS A L'ETUDE ET SUSCEPTIBLE DE MODIFICATION. IL NE PEUT SERVIR DE REFERENCE.

LES RECIPIENDAIRES DU PRESENT DOCUMENT SONT INVITES A PRESENTER, AVEC LEURS OBSERVATIONS, LA NOTIFICATION DES DROITS DE PROPRIETE DONT ILS AURAIENT EVENTUELLEMENT CONNAISSANCE ET A FOURNIR UNE DOCUMENTATION EXPLICATIVE.

THIS DOCUMENT IS STILL UNDER STUDY AND SUBJECT TO CHANGE. IT SHOULD NOT BE USED FOR REFERENCE PURPOSES.

RECIPIENTS OF THIS DOCUMENT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

Titre :

Title :

Maritime navigation and radiocommunication equipment and systems - Digital interfaces - Part 410: Multiple talker and multiple listeners - Ship systems interconnection - Transport profile requirements and basic transport profile

Introductory note

IEC 61162-4 Series specifies a communication protocol for use in integrated systems. It defines a ship wide and system level integration mechanism that complements communication solutions provided by other parts of the IEC 61162 series. It is also expected that the IEC 61162-4 Series will be used for data acquisition by higher level, non real-time and non-critical administrative workstations and personal computers. IEC 61162-4 Series has been developed as a network that can support a high number of nodes (several hundred if proper segmentation is used), with response times between 0.1 second and 1 second dependent on load. Ethernet and Internet protocols are employed at the transport level.

IEC 61162-4 has been divided into four different parts numbered IEC 61162-400, 401, 410 and 420.

ATTENTION **Parallel IEC CDV/CENELEC Enquiry)**	**ATTENTION** **CDV soumis en parallèle au vote (CEI) et à l'enquête (CENELEC)**

INTERNATIONAL ELECTROTECHNICAL COMMISSION

MARITIME NAVIGATION AND RADIOCOMMUNICATION EQUIPMENT AND SYSTEMS- DIGITAL INTERFACES-

Part 410: Multiple Talker and Multiple Listeners - Ship Systems Interconnection- Transport Profile Requirements and Basic Transport Profile

FOREWORD

1) The IEC (International Electrotechnical Commission) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of the IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, the IEC publishes International Standards. Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. The IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
2) The formal decisions or agreements of the IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested National Committees.
3) The documents produced have the form of recommendations for international use and are published in the form of standards, technical specifications, technical reports or guides and they are accepted by the National Committees in that sense.
4) In order to promote international unification, IEC National Committees undertake to apply IEC International Standards transparently to the maximum extent possible in their national and regional standards. Any divergence between the IEC Standard and the corresponding national or regional standard shall be clearly indicated in the latter.
5) The IEC provides no marking procedure to indicate its approval and cannot be rendered responsible for any equipment declared to be in conformity with one of its standards.
6) Attention is drawn to the possibility that some of the elements of this International Standard may be the subject of patent rights. The IEC shall not be held responsible for identifying any or all such patent rights.

This CDV for the International Standard IEC 61162-410 has been prepared by Technical Committee 80: Maritime Navigation and Radiocommunication Equipment and Systems.

[This CDV is proposed to cancel and replace the first edition of the CD]

The text of this standard is based on the following documents:

FDIS	Report on voting
XX/XX/FDIS	XX/XX/RVD

Full information on the voting for the approval of this standard can be found in the report on voting indicated in the above table.

This publication has been drafted in accordance with the ISO/IEC Directives, Part 3.

The committee has decided that the contents of this publication will remain unchanged until April 2003. At this date, the publication will be

- reconfirmed;
- withdrawn;

- replaced by a revised edition, or
- amended.

INTRODUCTION

The International Standard IEC 61162 has been prepared by Technical Committee 80: Maritime Navigation and Radiocommunication Equipment and Systems.

IEC 61162 is a four part standard which specifies four digital interfaces for applications in marine navigation, radio-communication and system integration.

The 4 parts are :

IEC 61162-1 Single Talker and Multiple Listeners

IEC 61162-2 Single Talker and Multiple Listeners - High Speed Transmission

IEC 61162-3 Multiple Talker and Multiple Listeners - Serial Data Instrument Network

IEC 61162-4 Multiple Talker and Multiple Listeners - Ship Systems Interconnection. This part is sub-divided into a number of individual standards with part numbers in the 400 series.

This part of the standard contains the specification of the requirements to an IEC 61162-4 transport profile (T-profile) and also the specification of one implementation, based on redundant Ethernet and Internet protocol functionality. The T-profile is the protocol transport mechanisms that offer simple message or byte stream transport services to the higher protocol layers (defined in other parts of the standard). In addition, the T-profile also offers services for time distribution and physical network management.

The use of Internet and Ethernet protocols offer low cost and high efficiency data transport in any kind of system. However, for safety related applications, certain measures have to be taken to avoid that particulars of office-quality and off-the-shelf technology create safety risks. This document specifies mechanisms by which a certain degree of quality of service can be guaranteed from these networks, including the provision of redundancy.

Other T-profile documents will be prepared with specifications of the same T-profile requirements over other transport protocols. This will be issued in the same number series as this standard (IEC 61162-41x).

CONTENTS

Page

1 Scope 7
1.1 T-profile 7
1.2 Contents of this fragment 7
2 Normative references 8
3 Definitions 9
4 Transport layer service specification 15
4.1 Introduction 15
4.2 T-profile network and host computer addressing 15
4.2.1 Introduction 15
4.2.2 Addressing scheme 15
4.2.3 Unique and singular hosts on the network 15
4.2.4 Non-segmented logical network 15
4.2.5 Size of network 16
4.3 Communication services requirements 16
4.3.1 Introduction 16
4.3.2 Connectivity and data transfer internally and externally 16
4.3.3 Service classes 16
4.4 Time services 19
4.4.1 General 19
4.4.2 Services 19
4.5 Management services 19
4.5.1 General 19
4.5.2 Services 20
4.6 Graceful degradation facility 20
5 Transport Layer Interface (TLI) 22
5.1 Scope 22
5.1.1 LNAC and TLI 22
5.2 General principles 23
5.2.1 TLI, TP Network service classes and connection points 23
5.2.2 Connection point capabilities 23
5.2.3 Attributes of TP network 25
5.2.4 Attributes of connection point 26
5.2.5 Connection point states 26
5.3 TLI service overview 29
5.3.1 General 29
5.3.2 TLI management services 29
5.3.3 Connection point management 31
5.3.4 Data management 33
6 T-profile for redundant Internet 36
6.1 Introduction 36
6.1.1 Rationale for the use of Internet protocols 36
6.1.2 Relationship to basic Internet protocols 36
6.1.3 Rationale for use of Ethernet 37
6.2 Transport profile architecture 37

6.2.1 General 37
6.2.2 Redundancy....... 38
6.2.3 Communication services 38
6.3 Message formats 39
6.3.1 General format 39
6.3.2 Special reliable link messages 40
6.3.3 Broadcast messages 40
6.4 Reliable stream protocol 41
6.4.1 Introduction 41
6.4.2 Connection management 41
6.4.3 Data transmission 44
6.4.4 Data reception 45
6.4.5 Link close 45
6.5 Reliable message service 45
6.5.1 Introduction 45
6.5.2 Data reception 45
6.5.3 Data transmission on the urgent link 46
6.6 Unreliable message service 46
6.6.1 Introduction 46
6.6.2 Connection management 46
6.6.3 Message length limitation....... 47
6.6.4 Data transmission....... 47
6.6.5 Data reception 48
6.7 Time services 48
6.7.1 Introduction 48
6.7.2 Time Management Architecture....... 48
6.7.3 Client interfaces....... 49
6.7.4 System integration requirements 49
6.8 Network management services 49
6.8.1 Introduction 49
6.8.2 Physical architecture....... 49
6.8.3 Required safety precautions....... 49
6.8.4 NMA requirements....... 50
6.8.5 Network element requirements....... 50
7 T-profile for non-redundant Internet....... 50
7.1 Introduction 50
7.2 Relationship to redundant Internet protocol 50
7.3 Redundancy 51
7.4 Reliable stream protocol 51
7.5 Reliable message service 51
7.6 Unreliable message service 51
7.7 Time services 51
7.8 Network management services 51
8 T-profile for wide area networks (MAU-LNA communication link) 51
8.1 Introduction 51
8.1.1 General architecture 52
8.1.2 TP-network and NNN code definitions 52
8.2 Relationship to non-redundant Internet T-profile....... 52

8.3 Reliable stream service 52
8.4 Reliable message service 53
8.5 Other services 53
Annex A (Informative) Typical software library structure 54
Annex B (Informative) Channel synchronisation example 55
Annex C (Informative) NTP and SNTP overview 56
Annex D (Informative) Overview of Internet Network Management Services 57

1 Scope

This part of IEC 61162-4 defines the general requirements of the T-profile and three implementations of the T-profile over the Internet V4 protocol suite. Part 400 of this standard defines the relationship between the different protocol levels (T-profile, A-profile and companion standards) and part 401 defines the A-profile, the immediate user of the protocol level defined in this part.

1.1 T-profile

The different components of the IEC 61162-4 standard are defined in IEC 61162-400. This clause gives a short overview of the function and structure of the T-profile.

The T-profile is the specification of the communication services and the communication protocols used by the LNA to implement the A-profile functionality. Basically, the T-profile consists of the following components:

a) A transport layer interface (TLI) definition that specifies the services and the semantics that shall be available to the application level of the LNA (and in some cases the MAU). This includes data transport as well as time and network management services. The TLI shall be general to all T-profiles and is defined in this fragment of the standard,

b) A T-profile protocol definition that shall specify how the services provided by the TLI and additional time distribution and physical network management services are implemented on the protocol level. This fragment contains a number of alternative T-profile protocol specifications using the Internet V4 series of standards. Additional parts of this standard will address other T-profiles based on other protocol families.

Note that the time distribution and network management functionality may or may not include specific TLI services. For some systems this functionality may be interfaced to directly by the underlying operating system. Note also that time distribution and network management are not strictly speaking transport related protocol functionality. However, the implementation of these services is normally dependent on the transport protocols in use and is, thus, placed in the T-profile part of the standard.

1.2 Contents of this fragment

The purpose of this document is to define and describe the services that shall be provided at the transport level interface in a way which is completely independent of the underlying network environment as well as defining one possible implementation of these services over the Internet V4 protocols. The separation of service and protocol definitions allows the specification of several different transport profiles, each one dedicated to a specific network environment, and to use the same transport service interface in all cases.

Clause 4 defines the transport level services and clause 5 describes the transport layer interface through which the services are offered. These clauses define the general, network independent services.

Clause 6 defines the transport profile architecture for redundant Ethernet and Internet protocols version 4 (IPV4). Clause 7 defines the architecture for a local area non-redundant Internet network. These clauses define two specific implementations of the T-profile services.

Clause 8 defines a simple MAU-LNA protocol for use over wide area network (WAN) TCP/IP links. This can be used to implement a WAN architecture for the overall system. The WAN architecture is not intended for integrated ship control systems, but can be used for remote test integration and remote maintenance and diagnostics. The WAN protocol can also be used to support MAUs that are located in other host computers than the LNA, but on one local network (conformance class 4).

2 Normative references

The following normative documents contain provisions which, through references in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All normative documents are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

Normative references are also defined in part 400 of this standard.

IEC 61162-4: (shorthand for all parts in the 400 series – to be published), Maritime navigation and radio communication equipment and systems - Digital interfaces - Part 4xx: Multiple talker and multiple listeners - Ship Systems Interconnection.

IEC 61162-400: (to be published), Maritime navigation and radio communication equipment and systems - Digital interfaces - Part 400: Multiple talker and multiple listeners - Ship Systems Interconnection – Introduction and General Principles.

IEEE 802.3: 1985, IEEE Standards for Local Area Networks: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications, IEEE, New York, New York, 1985, including all amendments.

ISO/IEC 9595:1998, Information technology -- Open Systems Interconnection -- Common management information service.

ISO/IEC 9596-1:1998, Information technology -- Open Systems Interconnection -- Common management information protocol -- Part 1: Specification and Part 2: Protocol Implementation Conformance Statement (PICS).

RFC 768: 1980, User Datagram Protocol (UDP), Internet Activities Board recommended standard.

RFC 793: 1981, Transmission Control Protocol (TCP), Internet Activities Board recommended standard.

RFC 826: 1982, Address Resolution Protocol (ARP), Internet Activities Board elective standard.

RFC 894: 1984, Internet Protocol on Ethernet Networks, Internet Activities Board elective standard.

RFC 1060: 1989, Assigned Numbers, Internet Activities Board required standard.

RFC 1042: 1988, a standard for the transmission of IP Datagrams over IEEE 802 Networks, to Internet Activities Board elective standard.

RFC 1112: 1989, Internet Group Multicast Protocol, Internet Activities Board recommended standard.

RFC 1157: 1990, Simple Network Management Protocol (SNMP).

RFC 1189: 1990, Common Management Information Services and Protocols for the Internet (CMOT and CMIP).

RFC 1213: 1991, Management Information Base for Network Management of TCP/IP-based Internets: MIB-II.

RFC 1305: 1992, Network Time Protocol, Version 3 -Specification and Implementation.

RFC 2030: 1996, Simple Network Time Protocol (SNTP),Version 4 for IPv4, IPv6 and OSI.

RFC 2500: 1999, Internet Official Protocol Standards- Internet Activities Board standard.

3 Definitions

This clause contains definitions that are applicable to this document alone. More general definitions can be found in part 400 of this standard.

For the purposes of this standard the following definitions apply:

Broadcast
See multicast.

CL - Connection-less
In this standard a connection less communication means that sender and receiver do not have to know each other. There is no association established between sender(s) and receiver(s) before messages are sent (see also peer-to-peer and client-server).

Client-server
A client-server communication link is established by the client after a server has allowed the connection attempt by the establishment of a listening connection point. The client is the active part while the server allows the connection (see also connection-less and peer-to-peer).

CMIP - Common Management Information Protocol
CMIP (Common Management Information Protocol) and CMIS (Common Management Information Services) are defined in [ISO/IEC 9595] and [ISO/IEC 9596].

CMIS - Common Management Information Services
See CMIP.

CMOT - CMIS/CMIP Over TCP/IP
CMOT (CMIS/CMIP Over TCP/IP) is an Internet proposed standard protocol. Its status is elective [RFC 1189].

CO – Connection Oriented
The opposite of CL (Connection less). A data exchange where an association between sender and receiver exists.

Connection point
An entity that can represent a communication link end point (for established connections) or a connection attempt between two host computers in some state. Connection point is also used

for connection less communication but in this case it represents just the local host computer's port to the network.

Ethernet.
References to Ethernet in this standard refers to a carrier sense, multiple access collision detect (CSMA/CD) local area network protocol standard as defined in [IEEE 802.3]. The medium access control (MAC) frame format shall use the Internet protocol type (0800) in the length/protocol field [RFC 894].

Any type of Ethernet can be used in systems compliant with this standard as long as they fulfil relevant technical requirements and the system integrator ensures compatibility between the integrated components. The most relevant technologies are:

- 10Base-5 - Thick coaxial, shared media bus. Can be used in environments that require noise immunity.
- 10Base-2 - Thin (RG58) coaxial, shared media bus. Can be used in environments with low noise immunity requirements.
- 10Base-T - Shielded or unshielded twisted pair. Used in conjunction with a repeating or switching hub.
- 10Base-F - Fibre optic media, as above, but for very high noise immunity applications.
- 100Base-TX - Shielded or unshielded twisted pair, as above. Note that this medium requires special precautions to extend network lengths beyond 200 m.
- 100Base-FX - Fibre optic, as previous, but for very high noise immunity.

The recommended solution is the use of non-duplex switching hubs with 10 and/or 100 megabit links to the host computers.

The system integrator must make sure that the relevant physical characteristics of the selected network solution satisfies the requirements defined by [IEEE 802.3] and other applicable standards.

ICMP - Internet Control Message Protocol
ICMP is an integral part of the Internet Protocols (see 0)

IGMP – Internet Group Management Protocol
IGMP is a protocol used between hosts and multicast routers on a single physical network to establish hosts' membership in particular multicast groups. Multicast routers use this information, in conjunction with a multicast routing protocol, to support IP multicast forwarding across the Internet. The IGMP standard is part of the Internet RFC system, but is currently not used in this standard.

IP - Internet protocol.
The term Internet protocol as used in this standard refers to [RFC 2500]. All required parts as defined by that document shall be implemented in an Internet compliant protocol stack. For the purposes of this standard it is also required that the system in question implements the ARP [RFC 826] address resolution protocol. Optionally, other address resolution schemes can be used at the discretion of the system integrator.

[RFC 2500] will give additional information as to which additional standards apply to conformant implementations of the Internet protocol.

[RFC 1918] lists a set of reserved private address spaces that can be used for a ship-board control network that shall not be connected to off-ship or other ship-board Internets. For the purposes of this standard it is recommended that the following network addresses are used:

172.16.0.1 to 172.16.255.254 (network mask 255.255.0.0 - class B).

IPV4 – Internet protocol version 4
This is the Internet protocol version used in this issue of the standard (see previous clause).

IPV6 – Internet protocol version 6
This is the proposed next generation Internet protocol. It is currently not sufficiently accepted to be viable as the basis of this standard, but future generations of this standard are expected to migrate to IPV6.

LAN – Local Area Network
A network within a delimited physical area. Typical for control system networks (see also WAN, MAN).

LNAC – LNA Communication module interface
A variant of TLI that is dedicated to MAU-LNA communication.

Loop-back interface
References to the loop-back interface in this standard refers to the loop-back function in the Internet protocol whereby an Internet message can be sent between two entities on the same network host. The complete class A network number 127 is assigned the “loop-back” function [RFC1060]. For the purposes of this standard, the address 127.0.0.1 will be used for the loop-back.

MA – Management agent
An SNMP agent located in a network element (NE).

MAN - Metropolitan Area Network
A network between LAN and WAN. The area the network covers is more than local, but not world wide.

MIB – Management Information Base
Basis for physical network management protocol [RFC 1213].

MTU - Maximum Transmission Unit
The largest message that can be transmitted over a given T-profile without fragmentation. For Ethernet, the MTU is 1500 octets.

Multicast
This standard specifies a transport mechanism by which any number of computer hosts can be reached with one transmitted message. This transport mechanism is referred to as multicast or broadcast. The T-profile specification defines the actual physical mechanism that is used to implement this service.

NE - Network element
Term used within SNMP to identify a managed entity in a network. This is typically a host computer with its protocol entities, a router, a gateway or a switch.

NMA _ Network Management Applications
Applications within a NMS.

NMS – Network Management Station
A host computer, with software implementing one or more NMA, that is responsible for physical network monitoring (see MIB and SNMP).

NNN - Network node number
An identification of one node on a TP network. This is typically the Internet host number (excluding the network part of the address) for IPV4 networks. The NNN shall be unique and unambiguous for one node, even if this node is on a redundant network with two different network addresses. The NNN is a 32 bit unsigned integer.

NTP - Network Time Protocol
All references to NTP in this standard address the Internet time distribution and synchronisation protocol described in [RFC 1305].

OSI – Open System Interconnect
This term is defined in [IEC 61162-400].

Peer-to-peer
A two-way communication link that can be established by both sides concurrently. The T-profile shall make sure that only one link is established and that the link has the same identification on both sides (see client-server and connection less).

Primary network
A redundant network has two independent transport paths between any two nodes (primary and secondary network). The primary network is the transport path defined as the one that is normally used when connections are established between the nodes, and that normally is to be used if non-redundant nodes are attached to the network. However, the T profile shall be able to use the secondary network in the same manner as the primary (i.e., accept incoming connects that arrive on the secondary network alone). The T-profile may process request on the secondary network after requests on the primary.

Priority
This standard uses the term priority to assign importance to messages or message streams transported through the communication network. Three levels of priority are defined: Urgent, normal and low.

The priority shall (as a minimum) act as a rudimentary bandwidth allocation scheme where high priority (importance) data is serviced before lower priority data.

QoS – Quality of Service
To what degree a certain protocol (typically T-profile) supplies certain services to the higher levels [IEC 61162-400].

Redundant IPV4
The reference to a redundant Internet shall mean a T-profile as defined in clause 6 of this standard. Where the standard redundant IPV4 network is used (`LNADR_IPV4R`), the following addresses shall be used:

172.16.0.1 to 172.16.255.254 (network mask 255.255.0.0 - “primary”).

172.17.0.1 to 172.17.255.254 (network mask 255.255.0.0 - “secondary”).

Other addresses can be used when the system integrator defines new T-profile networks for use in a specific implementation (`LNADR_IPV4RE`).

Both networks shall be of the same class (class B for the standard network) and each network node shall have the same node address on the primary and secondary networks.

For all IP networks the following port numbers shall be used:

Table 1 - Internet IPV4 protocol port numbers

Port type	Protocol	Port
LNA-LNA listening port (normal)	TCP	23305
LNA-LNA listening port (urgent)	TCP	23306
LNA-MAU listening port (normal)	TCP	23303
LNA-MAU listening port (urgent)	TCP	23304
MAU-MAU stream listening ports	TCP	22307-
ABC0 (Standard broadcast)	UDP	23307/23308
ABC1 – ABC5	UDP	23309/10 – 23317/18

NOTE 1 - The port numbers are different from the ones used in version 3 of the protocol. This is done to support the co-existence of V3 and V4 networks.

NOTE 2 – The stream ports will be selected by trying different port numbers in sequence, starting with the one specified.

Two port numbers are defined for each UDP broadcast port to enable a host computer to use different listening and sending ports. The first number (odd numbered) shall be the listening port. The immediately following even number is the output port.

Likewise, two port numbers are used for the message ports to support differentiation between normal and urgent messages. Implementations that do not support differentiation shall use the normal port (valid for MAU-LNA link only).

The same ports shall be used for all IP type networks and on both the primary and secondary side of a redundant network.

Secondary network
See primary network (clause 0). The secondary network represents the second transport path.

SNMP - Simple Network Management Protocol
Internet protocol used for physical network management [RFC 1157].

TCP/IP - Transmission Control Protocol/Internet Protocol
References to TCP/IP in this standard refer to the reliable and stream oriented protocol defined by [RFC 793]. A system implementing TCP/IP shall also implement the Internet protocol as described above.

TPN – TP-network
A T-profile network (TP-network) is a definition of a physical network architecture with all relevant address and network parameters. This fragment of the standard defines a number of different TP-networks based on the IPV4 family of protocols.

UDP - User Datagram Protocol
References to UDP in this standard refer to the unreliable message based protocol defined by [RFC 768]. A system implementing UDP shall also implement the Internet protocol as described above.

WAN – Wide Area Network
A network that, in principle, can span the entire world. Typical example is the Internet.

4 Transport layer service specification

4.1 Introduction

The transport layer will provide the upper layers (the MAUs and the A-profile) with functions for the following services:

- Communication services,
- Time services,
- Network management services,
- Graceful degradation facility.

Communication and graceful degradation services are provided by the transport layer while time and management services typically involve the transport layer and/or other entities (upper layers, resources depending on actual implementation).

4.2 T-profile network and host computer addressing

4.2.1 Introduction

This clause defines requirements to the T-profile network and host computer addressing. The described addressing scheme shall ensure easy portability for higher level entities between different T-profiles.

4.2.2 Addressing scheme

Host computers that communicates via a specific T-profile use a two-level addressing scheme:

- Each T-profile shall as a minimum, define one T-profile network (TP-network) that specifies the physical architecture of the network. This network is logically a linear network without any further sub-divisions or segmentation. Some T-profiles may define more than one TP network to sub-dive the different functions supplied by the T-profile into separate modules.

  NOTE – A typical example is a redundant Internet profile that sub-divides the TP network into one for MAUs (supporting only stream communication) and one for the LNA (supporting message based communication).

- Each host computer has a unique and singular network node number (NNN). This number when referenced to a specific TP network shall be sufficient to address the host computer by the T-profile implementation.

4.2.3 Unique and singular hosts on the network

The T-profile shall provide an addressing mechanism that makes each host computer on a T-profile network uniquely and singularly identifiable by a network node number (NNN). This applies for all types of T-profile networks, for those that are non-redundant as well as redundant. One NNN shall map to one and only one possible host computer.

NOTE - For redundant networks the single NNN must normally be convertible to two physical network addresses, one for each of the active links.

4.2.4 Non-segmented logical network

The addressing scheme defined here creates a non-segmented logical network on which the host computers reside. However, a T-profile network may optionally provide physical segmentation between groups of host computers in the NNN address space. This standard does not specify how this can or should be done. This standard requires that all nodes on the T-profile network, independent of physical segmentation, shall be able to establish communication links between themselves and exchange data over these links.

NOTE - A T-profile for the wide area Internet may need to handle broadcasts specially (e.g., by using a multicast protocol, e.g., IGMP) to enable general MAU name look-up over several networks.

4.2.5 Size of network

The NNN is an unsigned integer of at most 32 bits in length. Physical addressing schemes that require more bit-space must define attributes in the T-profile network description that supply the extra information so that a shorter NNN can be mapped to the correct address.

NOTE - IPV6 uses a 128 bit long network address. To make host computers addressable by 32 bits, the remaining 96 bits must be determined from the T-profile network attributes. The same applies to redundant networks, e.g., for IPV4, where each host computer typically has two network addresses.

4.3 Communication services requirements

4.3.1 Introduction

This clause defines the general attributes of the communication services that shall be made available through the T-profile.

4.3.2 Connectivity and data transfer internally and externally

The T-profile shall provide the following general services:

- The transfer of data between any nodes attached to the embedded system level network.
- Allow connection to one or more other networks than the embedded system level network.

  NOTE - The concept of defining different logical T-profile networks allows several T-profiles to coexist on the same physical network. In this case, the *other* network may be another T-profile. More commonly, the other network is some form of administrative network accessed via a separate network interface card (NIC).

Different service classes are defined in this standard for data transfer on the embedded system network. Data transfer out of the embedded system network is not covered by this standard.

4.3.3 Service classes

4.3.3.1 Overview

This clause defines the service classes that shall be supported. Table 2 summarises the service classes and their main characteristics.

Table 2 - Service classes

Communication	Connection	Data	Priority	Relation	Instances
LNA-LNA data	CO	Msg	U and N	peer to peer	One
LNA-LNA MC in	CL	Short msg	U and/or N	multi-cast listen	6 MC ports
LNA-LNA MC out	CL	Short msg	U and/or N	multi-cast send	6 MC ports
MAU-MAU stream N	CO	Stream	N	client/server	Any
MAU-MAU stream L	CO	Stream	L	client/server	Any
MAU-LNA data	CO	Msg	U and N	client/server	One

The first column specifies the service class and by that, its normal area of use. The second column specifies whether the communication uses a connection oriented (CO) or connection-less service. The third column specifies whether data is sent as messages (with message boundaries) or as streams. The next column specifies the priority classes that shall be supported for this service class. The next to final column specifies the relationship between host computers using the service.

The final column specifies the various instances that are supported for each communication class. For LNA-LNA and MAU-LNA links this is just one: The one standard path to the one listening LNA. For multi-cast services, the A-profile requires 6 predefined addresses, one for name look up and five for anonymous broadcast. For MAU-MAU link the TP network should support a reasonable number of CP to allow possibly multiple MAUs on one host computer to have multiple connections to other MAUs.

The communication service classes are:

- LNA-LNA data: This is a communication link providing reliable bi-directional data transfer between any two LNAs (communication initiated by any party – peer to peer).
- LNA-LNA MC: This is a broad- or multi-cast service that provides name look-up, watchdog and anonymous broadcast services between LNAs. This is also essentially a peer to peer communication class, although connection less. The service is divided into a transmitting and a receiving component.
- MAU-MAU stream: This is a service providing bulk data transfer between MAUs. The communication is initiated by a client connecting to a waiting server. There are two stream classes, one for normal and one for low priority (priority is associated with the link rather than with the data transported). The service is bi-directional.
- MAU-LNA data: This is a service (defined in the A-profile) providing communication between a MAU and its LNA. This is also a client-server type communication. The service is bi-directional.

4.3.3.2 Implementation of multiple service classes

One implementation of a T-profile may provide one or more of the above service classes, dependent on the required use. The minimum requirements are listed in the below table.

Table 3 – T-profile service implementation

TP Implementation for	Required service classes
LNA-LNA	LNA-LNA data, LNA-LNA MC
MAU-MAU	MAU-MAU stream N, MAU-MAU stream L
MAU-LNA	MAU-LNA data

NOTE – Although all these services may be implemented by one T-profile, the implementation may offer the services to the users as a different TP-network. Typically, the three groups of services will be offered as three distinct TP networks.

4.3.3.3 Message or stream type communication

The following communication services shall be provided to the upper layers:

- Msg: Reliable message transfer. The transfer mode will be a unicast (point to point) transfer mode. Message length may be limited by T-profile implementation. In particular, urgent messages may be limited to the MTU of the T-profile.
- Stream: Reliable stream transfer of a unicast (point to point) transfer mode. Data is transmitted between sender and receiver without regard to message boundaries.
- Short msg: Unreliable short message transfer. The transfer mode will be a multicast (point to multipoint) or broadcast transfer mode. The message length is typically limited by the MTU of the T-profile (and, hence, is “short”).

4.3.3.4 Connection modes and flow control

A reliable transfer will be achieved through a Connection Oriented (CO) mode protocol while an unreliable transfer will be achieved through a Connection-Less (CL) mode protocol.

A CO mode protocol provides for full duplex reliable (with guarantee of delivery to the destination), sequential and without duplication end to end data transfers by using sequence numbers, acknowledgements to indicate correct delivery of the data, and retransmissions in case of error (data loss, data damage).

A flow control mechanism shall be provided for CO services to avoid a receiver being congested by too many arriving messages.

A mechanism should be provided to detect that either side of an established connection has ceased to function (remote network node damage). This mechanism will in any case be augmented by a similar service in the A-profile so it will not be of critical importance.

A CL mode protocol provides for unreliable (without guarantee of delivery to the destination(s)) end to end message type data transfers. No notification is generated for either loss, duplication or out of sequence delivery of data. Damaged data need not be retransmitted, but shall in no case be delivered to the receiver (i.e., discard messages with bad checksums). The service is specified to be unidirectional (i.e., a listening service must be used together with a sending service).

Unreliable messages shall be short messages. It will be the responsibility of the upper layers to ensure that the length of the messages provided at the transport layer is within the permitted maximum value (medium MTU).

Reliable streams / long messages shall be segmented and reassembled at the transport layer when necessary. Short messages (typically of urgent priority) should, if possible, be transmitted in one physical layer message.

4.3.3.5 Priority

In order to support the time dependency of the transmitted data (time critical or not) and to allow the intermixing of data flows of different significance on the network, importance levels shall be defined as follows:

- Urgent.
- Normal.
- Low.

The implementation of these priorities will be T-profile dependent. In particular, there may not be a well defined relationship between priority levels for CL and CO services as the CL and CO mode transfers may use different protocols. The T-profile shall in any case provide a best effort implementation of priorities.

The system integrator must verify that the implemented priority mechanism is suitable for the use to which the T-profile is being put.

4.3.3.6 Logical communication paths and node relationships

A logical communication path results from the association between two (point to point communication) or more (point to multipoint communication) transport level end points.

Dependent of communication class, different principles apply for the establishment of logical paths:

- Peer to peer: Communication between two LNAs can be established by any part at any time. The T-profile implementation must make sure that concurrent connection establishments result in only one logical path.

  NOTE - To facilitate a similar handling of all types of communication links, also the peer-to-peer service will be established in the same way as a client/server connection.

- Client-server: Communication between LNA and MAU and between MAUs is established by one of the parties acting as client (doing the connection request) and the other acting as server (waiting for connection attempt). This always results in one unambiguous communication path.

- Multicast: This communication is connection-less and does not use logical paths.

4.4 Time services

4.4.1 General

A common time reference shall be available in the system level network in order to timestamp all time dependent data transmitted by the nodes and to estimate the validity of the received data.

The time shall be generated and maintained at the network level at least at one dedicated time generator/server. The time distribution service shall allow the use of multiple (redundant) servers.

A time distribution service shall be provided to update and to synchronise the local clock available at each attached node to the global clock available at the time generator/server.

In addition, the time shall be maintained at each node in order to make available to their local user(s) a reference time even in case of failure of the dedicated common reference time generator/server(s).

The T-profile specification shall minimise synchronisation jitters by, e.g., using urgent priorities for time related messages. Other means can also be employed. The worst case jitters shall be specified.

4.4.2 Services

The time service shall maintain a synchronised time over the complete T-profile network. No particular services are provided to the applications. It is assumed that the time related services are made available through the operating system interface.

4.5 Management services

4.5.1 General

The management services shall enable monitoring and control of the system level network operation.

It requires monitoring and control of the different network elements: Attached nodes and intelligent network equipment (e.g., switch, router or gateway).

Each one of these network elements shall maintain management information providing information (e.g., state, thresholds, counters, action results) on the local resources (software and hardware).

The management information involves standard management information (information dedicated to standard communication protocols), proprietary management information (information dedicated to manufacturer’s equipment) and if necessary specific management information (information dedicated to specific system requirements).

Network configuration, fault and performance management functions are achieved from a dedicated network management station (NMS) providing the network maintainer with the current status of the network as an aggregate of the different management information distributed in the network.

The NMS accesses to the management information, local to each network element, through a standardised network management protocol, generally at the OSI application layer, while local management functions on a node rely on specific management services depending on the node implementation (operating system and protocol stack features).

4.5.2 Services

The following management services shall be provided:

- Monitoring of a node attached to the network (load and fault conditions).
- Network monitoring (load and fault conditions).

The following services are optional, but may be prohibited due to safety considerations:

- Remedial actions.
- (Re)configuration.

Each node shall detect internal errors and make available detailed status information to the local upper layers and to the remote network management station(s) (NMS).

The NMS shall build up the overall network status (list of network elements in operation) by:

- Receiving reports of changes from the nodes and intelligent network equipment (e.g., activation, deactivation and fault reports).
- Requesting, with certain frequency, status, events, action results and statistical information from the same elements (if the information is not automatically broadcast).

The NMS shall monitor the network traffic and build traffic statistics.

Local node management shall be sufficient to enable a user to monitor the condition of its communication to the network and take remedial actions such as configuration change, network disconnection or test initialisation. Any such actions shall only be available when the system is in a safe state (e.g., ship at berth) and in a way that cannot endanger the ship's safety.

The information returned from the network monitoring activity shall be sufficient to enable remedial actions to be taken such as network equipment configuration change, disconnection or reset.

The NMS must normally be duplicated for safety purposes.

4.6 Graceful degradation facility

The system level network shall be designed with a capacity for graceful degradation in order to retain maximum performance from the remaining available resources in case of a failure.

Redundancy should be used for safety critical parts of the system in order to provide services and performances even after a single failure.

In case of a number of faults exceeding its resilience, the network shall still provide its services to the maximum possible number of nodes considering that time budget performances may not be guaranteed and that a part of the network may not be available. The network is then in a degraded condition; the gravity of the degradation depends on the types of faults.

When operation of a resource is degraded or fails, an indication shall be provided to the NMS and to the A-profile where possible and relevant (see previous clause).

Graceful degradation shall be automatic and transparent to the users at least on the occurrence of these events:

- Failure of a single cable or single network interface hardware (no loss of functionality).
- Fault, insertion or shutdown of one or more nodes (loss of function of that node). The node may have to be duplicated for safety purposes.
- Loss of all the network equipment within a single compartment (due to fire or flood – loss of functionality located in compartment).

The implementation of graceful degradation is partly dependent on physical architecture.

5 Transport Layer Interface (TLI)

5.1 Scope

The A-profile shall get access to the transport layer services through a transport layer interface (TLI). The TLI shall make the actual transport profile transparent to the upper layers.

The transport layer interface shall provide the upper layers with a set of standard transport service primitives independent of the network environment while its implementation will depend on the actual environment (e.g., transport profile, operating system and local resources). The TLI will normally not supply time and network management services.

The transport layer interface is illustrated in the following figure.

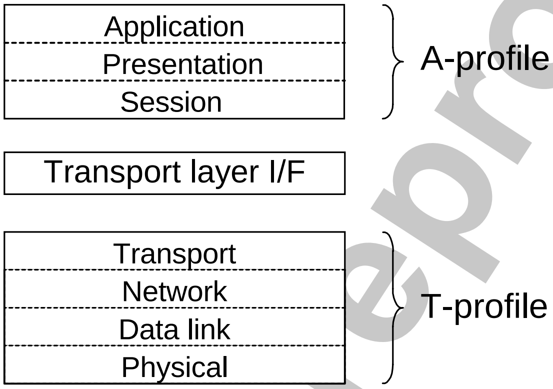


Figure 1 - Transport layer interface

The purpose of the TLI is to define a standard interface between the A-profile and the T-profile. The A-profile specification (part 401 of this standard) depends on the TLI service descriptions contained in this clause.

The service specification will not require any particular language binding or any particular modularization of the services into specific subroutine calls. This is left to the implementation. However, the service specification will be made in such a way so that it can be translated one to one into a normal procedural language.

Annex A describes a typical software library structure for the TLI and associated modules.

5.1.1 LNAC and TLI

A special variant of the TLI is the LNA Communication module (LNAC). This module is used to set up communication between LNA and MAU. For MAU-LNA communication over a normal network communication path (e.g., TCP/IP) it is normally most convenient to implement the LNAC as a normal TLI with the same services and a specialised TP-network. For IPC type communication between LNA and MAU, it may be more convenient to implement the LNAC in some other way.

This standard defines requirements to the TLI and associated T-profiles. These requirements can optionally be used also for the LNAC. An Internet LNAC profile is defined in clause 8.

5.2 General principles

5.2.1 TLI, TP Network service classes and connection points

To have a general communication mechanism, the concept of a connection point is used. The connection point is a communication link end point and is directly associated with the TLI. However, the connection point derives address information and capabilities from the TP network and the service class it belongs to. The relationship between the TLI, TP networks, service classes, instances and connection points are illustrated in the below ER diagram.

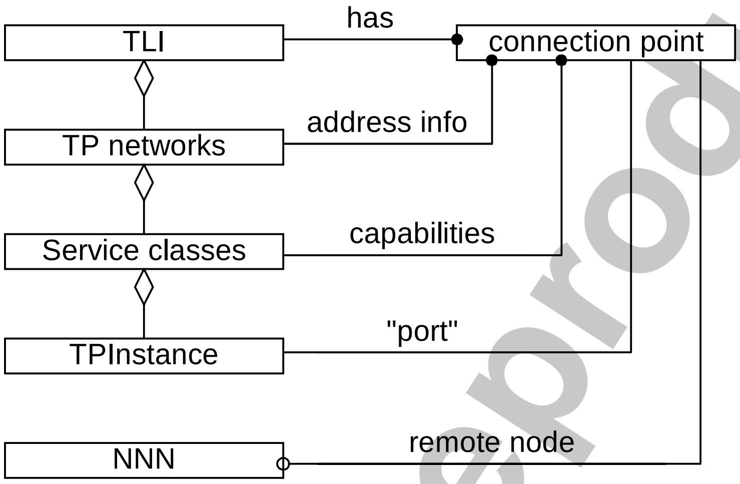


Figure 2 - Relationship between TLI, TP networks and ports

One TLI shall support the use of a number of different TP networks (minimum one, usually two or more, e.g., the LNA needs one network for the LNA-LNA communication and will typically use one or two more TP networks for LNA-MAU connections). Each TP network supports one or more service classes, dependent on network.

NOTE - Even though one TP network specified in this standard can support several service classes, the implementation may segment the one TP network into different modules – normally based on service class. This is convenient for modularization, but may mean that the programmer has to select between different TP network modules to find the service class he or she is looking for.

For each service class, it will normally be possible to create a number of communication service instances. The services may be just different instances of the service (e.g., concurrent MAU-MAU streams between two host computers) or some meaning may be pre-allocated to the instances (e.g., a pre-defined multi-cast group). One CP can be created for each instance. Finally, each instance of a connection point will normally also have a remote node associated with it. This is, however, not the case for connection less services.

5.2.2 Connection point capabilities

The capabilities of a connection point (CP) are derived from one of the service classes supported by the network as defined in Table 2. However, to support client-server type link establishment, it is necessary to distinguish between listening and connecting CPs. Thus, the complete list of CP types is as follows:

Table 4 – Connection Point Types

CP type	Priority	Type	Relation	Parameters
CPL_LNAMSG	n/a	Listen for incoming LNA-LNA	peer to peer	None
CP_LNAMSG	n/a	End point LNA-LNA	peer to peer	Remote NNN
CP_LNAMC_IN	n/a	Short message input	connection less	ABC port

CP type	Priority	Type	Relation	Parameters
CP_LNAMC_OUT	n/a	Short message output	connection less	ABC port
CPL_STREAM	L or N	Listen for incoming stream	client/server	Port
CP_STREAM	L or N	End point reliable stream	client/server	Remote NNN, port
CPL_MAULNA	n/a	Listen for incoming MAU-LNA	client/server	None
CP_MAULNA	n/a	End point MAU-LNA	client/server	Remote NNN

The LNA-LNA connection point is listed as a peer-to-peer connection, but this standard requires that it is implemented as a client/server port by the TLI (i.e., have a listening variant). The difference from normal client/server links, e.g., stream links, is that any concurrent connection attempts from the two peers automatically are resolved into one CP at each side.

The connection-less CP has one listening and one sending variant. This is to simplify the handling of multi-casts where the sender is concurrently listening in on the same multi-cast group as it is transmitting to. Outgoing messages shall be copied to the incoming local CP (if open) by the T-profile.

Message connection points accept incoming and outgoing messages of any priority. Stream connection points have an inherent priority that is established when the port is created.

5.2.2.1 Client-server connection establishment

A client-server CP requires that one host computer acts as a server and another as client. The client (the connecting) host computer will not be allowed to establish a connection before the server has established a listening port. The server will only establish a connection as a result of the client doing a connection attempt.

The T-profile shall accept that two “identical” CP pairs are established when the two host computers concurrently establish a listening and a connection port each.

NOTE – This is contrary to what is required of the peer to peer CP. The rationale is that client-server links are established as a result of a temporary need to exchange data (bulk data between MAUs). The situation outlined above can occur if both MAUs want to transmit data concurrently.

5.2.2.2 Peer-to-peer connection establishment

Two peer-to-peer CP pairs can be requested concurrently when two host computers want to communicate. The T-profile shall make sure that only one CP pair results from this process. In this standard only the LNA-LNA connection is of this type.

NOTE – This situation is typical of system start-up where many LNAs establish connections to each other. As there is no client-server relationship between LNAs, this requirement makes sure that each remote LNA is associated with exactly one CP.

To facilitate homogenous processing of connection points in the A-profile, the T-profile (and TLI) shall provide a listening as well as a connecting variant of the peer to peer CP. No incoming connection attempts shall be served before the listening CP has been established.

NOTE – This requirement simplifies the use of protocol software that explicitly use a client-server model, e.g., sockets for TCP/IP. By having this requirement, the T-profile will get an explicit instruction to create a listening port when the LNA is ready to process incoming connection attempts.

5.2.2.3 Connection-less connection points

Connection-less communication (multi-cast) shall also use connection points. Although there is no connection between host computers in this case, the local T-profile shall establish a connection point to act as reference for incoming and outgoing messages.

Each connection point shall only be able to handle either incoming or outgoing messages. Two CP are necessary to both send and receive multi-casts.

NOTE – If the T-profile uses multi-cast instead of broadcast, it must also automatically establish the multi-cast groups based on parameters in the TP-network object and probably when the TP-network is created. The user (LNA) will only request the establishment of a multi-cast CP and assume that it will work as specified.

5.2.3 Attributes of TP network

Each TLI TP network will have a number of attributes that characterise its functionality. The minimum set of attributes is defined in the below table.

Table 5 - TPN Attributes

Attribute code	Type	Description
TPA_NETWORK	int	Code (type) of TP-Network
TPA_SERVICES	word32_m	Bitmap for available services
TPA_QOS	word32_m	Bit-map defining quality attributes
TPA_MTU	int	MTU of T-profile (Zero if no limit)
TPA_STATE	int	State of TP (fully/reduced) operation
TPA_ADDR	address_m	Network address
TPA_LNNN	word32_m	Local network node number

5.2.3.1 TP-network address formats

The TP network address format will be dependent on the T-profile in use. The TLI shall adopt a T-profile independent address format as specified below:

a) One `word16_m` entity specifying the type of network address (e.g., Internet IPV4, redundant Internet IPV4/Ethernet, redundant IPV6/Ethernet, redundant ARCNET or other). Table 6 and details in each T-profile definition will determine the identifier.

b) One `word8_m` entity giving the length of the following address field.

c) A sequence of maximum 48 `word8_m` entities containing the actual network address as required by the T-profile in use.

This is a record with the following format string [IEC 61162-401]: “`(w16[w8:48]w8)`”.

5.2.3.2 Defined TP-network code ranges

The following table lists the TP networks that are defined by this fragment of the standard with required TP network and NNN address formats.

Table 6 - TP network and NNN types

Code	Values	Description	Notes
ANYADR_ANY	0	Undefined	For user specified addresses.
MAUADR_IPC	1	IPC type MAU-LNA link.	Not defined in this standard. Reserved for LNAC.
MAUADR_TCP	2	IPV4 TCP MAU-LNA link.	See clause 8.
MAUADR_TCPV6	3	IPV6 TCP MAU-LNA link.	Reserved for a future standard
LNADR_ARCNET	10 to 15	ARCNET range addresses	Reserved for a future standard.
LNADR_IPV4	120 to 139	IPV4 address formats.	See clauses 6 and 7.
LNADR_IPV6	140 to 159	IPV6 address formats.	Reserved for a future standard.

NOTE – Even with the same TP-network code, different address formats can be supported, e.g., for the user defined network and for the IPC network. In these cases, the address field must contain enough information to distinguish between the various address forms.

More address codes can be defined in future parts of this standard.

5.2.3.3 Compatibility between redundant and none-redundant TP networks

This standard allows the T-profile to define compatible redundant and non-redundant TP-networks. In these cases, hosts on a redundant TP networks shall be able to connect to hosts on the corresponding, non-redundant TP network and vice versa.

NOTE – This is useful in a scenario where some non-critical and low cost host computers shall be interfaced to the safety-related system network for, e.g., data acquisition or periodic maintenance.

It is allowed to let one (basically redundant) TP-network accept both types of incoming connections. The T-profile is required to determine what kind of network the remote connection host computer resides on and report this to the local application.

NOTE – An alternative is to let the local host computer define a TP-network for both types of hosts that it will connect to or allow connections from. Typically the local host computer will define one redundant and one non-redundant TP-network.

5.2.4 Attributes of connection point

Each connection point will have a number of attributes that partly are set permanently during the creation of the CP and partly represent the state of the port. All attributes should be retrievable through the TLI calls. The minimum set of attributes is listed in the below table.

Table 7 - CP Attributes

Attribute code	Type	Description
CPA_NETWORK	int	Identity of TP-network
CPA_SERVICES	int	Code for selected service
CPA_PRIO	word32_m	Bit-map showing priorities available
CPA_STATE	int	State of CP
CPA_RNNN	word32_m	Remote network node number
CPA_ADDR	address_m	Local address (including, e.g., port number)

In addition it will maintain a dynamic state vector keeping information on queue status and traffic statistics data.

5.2.5 Connection point states

The following state diagrams use the same set of states:

a) TP_DEF: The CP attributes are defined and made known to the TLI. The TLI is waiting for a connection request, either from local or remote host computer.

b) TP_REQ: The application has requested that the CP shall be made ready for communication. This state is used for a connecting CP that needs to establish a connection with another host computer’s listening CP.

c) TP_CON: This state is used for CP that are in a communicable state, i.e., messages can be sent and received or connecting CP can be accepted.

d) TP_ERROR: Errors, except time-out and forced close, make a CP go to this state. The CP will have to be reset (removed) before it can be used anew.

e) TP_CLOSE: A close request will normally make the CP go into this state (except for ports in the TP_REQ state). This state allows queued messages to be sent before closing the connection.

NOTE 1 - Common for most state transition diagrams is that they contain an explicit created and error state. These states are used so that all external events (close, error) cause a transition to a known state in a still defined object. The object can only be deleted through an explicit call from the higher level program. The exception is the CP that is derived from a listening CP (the “spawned” CP).

NOTE 2 - The existence of the created state also makes it convenient to let the close call move the CP to created instead of deleting it. This is not strictly speaking necessary, but makes the diagrams more consistent.

5.2.5.1 Listening connection points

The following figure defines the states of a listening connection point.

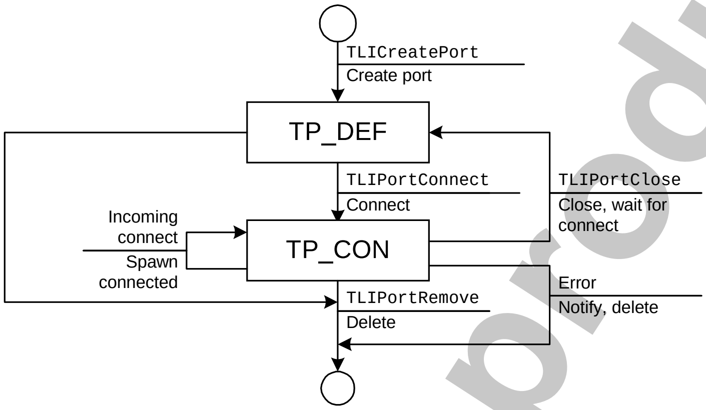


Figure 3 - Listening connection point states

NOTE - Note that this connection point is almost state-less and that it “spawns” new connection points during connection attempts.

5.2.5.2 Connecting connection point

The following figure defines the states of the connecting connection points.

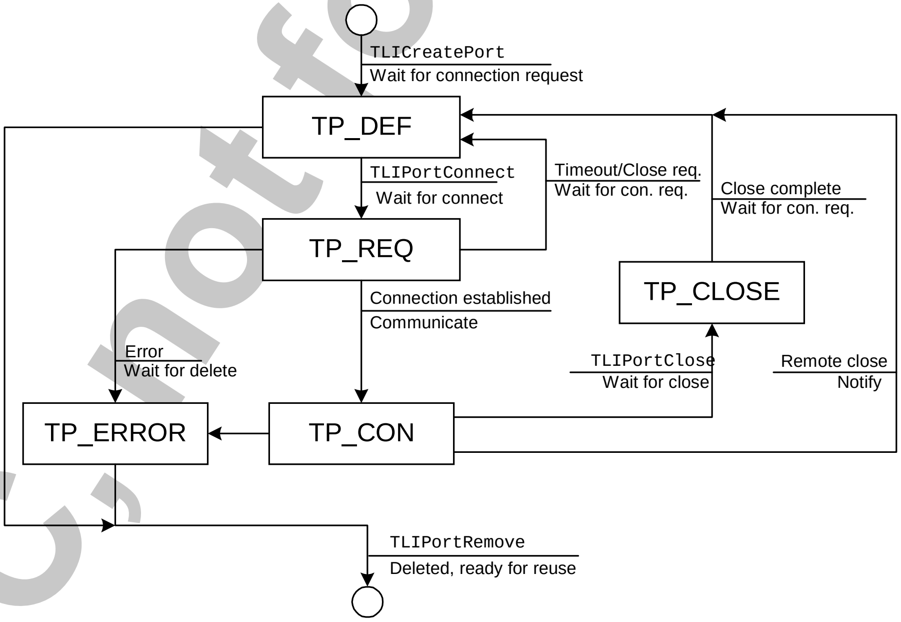


Figure 4 - States of connecting connection points

NOTE - The connection point must remember its attributes even if the connection is broken or if the CP is closed by the application. This makes it easier to restore the connection afterwards. Note also that errors that inhibit further connection attempts put the connection point into an error state.

5.2.5.3 Accepting connection point

The following figure illustrates the states of a CP spawned from a listening connection point. This CP does not use a “created” state as it is derived from the listening CP. The notification given to the higher level application in conjunction with the close of the CP, shall be given while the CP is still in existence. The removal can take place after the notification has been given and processed.

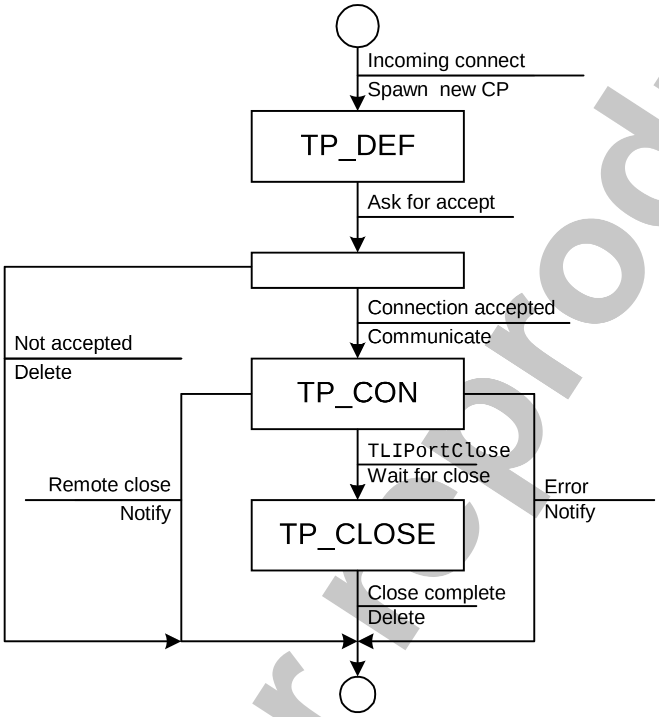


Figure 5 - States of accepting connection point

5.2.5.4 States of connection-less connection points

The states of connection-less CP are shown in the figure below. This is similar to the listening CP since this CP is essentially stateless.

Messages can be sent and received in the connected state. The connected state is reached immediately after requesting a port connection.

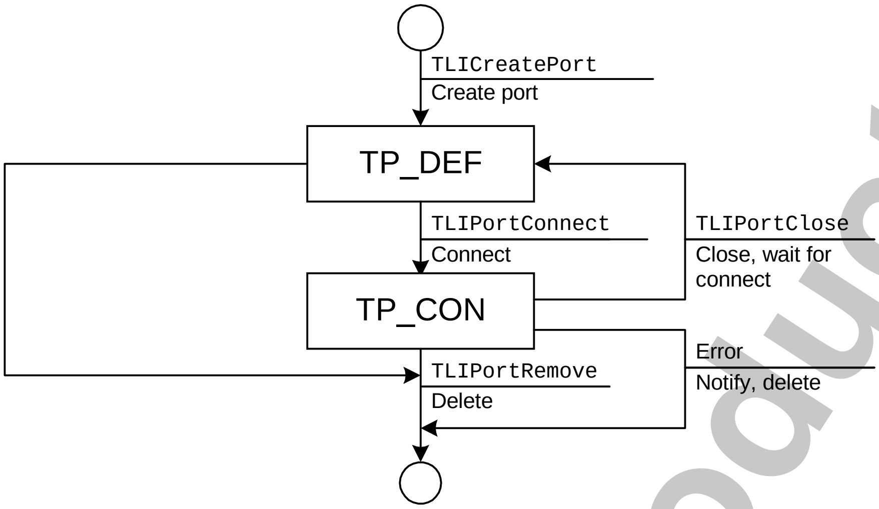


Figure 6 - Connection less connection point states

5.3 TLI service overview

The following clauses describe the required TLI service primitives. The services described in these clauses are required, but the actual TLI may be implemented by using other (e.g., other functions or names, fewer or more) service primitives.

NOTE - The specification of the A-profile is partly modelled on the availability of the services below. It is recommended that an actual implementation of a TLI use these services as they are specified. Otherwise, a rewrite of the A-profile implementation may be required.

5.3.1 General

5.3.1.1 Programming paradigm

The TLI described here is based on an event oriented principle where no service invocation blocks program execution (no waiting for external events) and where external events are delivered asynchronously to the callers event handler. The service description format follows that specified in the annexes of IEC 61162-400.

5.3.1.2 Structural assumption

It is assumed that the TLI itself is a T-profile independent library that manages the interface between the T-profile and the A-profile. The TLI will thus have to be enhanced with TP network specific modules that deliver the actual functionality. This is reflected in the existence of the “add TP network” function.

5.3.2 TLI management services

5.3.2.1 TLI Initialisation (TLIInitialise)

Request TLI to establish necessary connections to operating system and enable the addition of TP networks. Initialise buffer management and other TP-network independent parts of the TLI.

Pre-condition: None.

Post-condition: TLI ready for use. No TP-networks defined.

Input parameters:

- Upper limit size for buffer pool.

- Other implementation dependent configuration limits, e.g., maximum number of CP.

NOTE – To enable the creation of a TLI without using dynamic memory management it is suggested that the upper limit size is transferred as a memory block of a suitable size. This memory block can then be used by the TLI for its buffer and object store requirements.

Output parameters: Status of call.

Call-back: This call shall block caller until complete. No call back or exception handler is used later. Any later exceptions will only occur relative to TP networks or connection points.

Errors: Various operating system errors:

- Out of memory (application can be asked to give more memory or to close ports)
- Various internal failures (bad pointers, inconsistencies)

5.3.2.2 Adding a TP network (TLIAddTPN)

This service adds a new TP network to the list of supported TP networks.

Pre-condition: TLI initialised.

Post-condition: New TP network can be used.

Input parameters:

Descriptor structure that allows TLI to dispatch calls to and from the new TP network modules.

Output parameters: TP network identifier.

Call-back: TP network related exceptions:

- Network warnings (excessive errors, reduction of redundancy)
- Complete TP network failure (all CP related to TPN back to defined state).

Errors: Various operating system errors (e.g., out of memory).

NOTE – This service can be implemented by the TLI TP-independent service layer as a way to let new TP-networks specify their services that are available. This requires that the user layer calls some kind of initialisation routine in the TP-network specific module.

5.3.2.3 Change or inspect TPN attributes (TLITPNAttribute)

This service can be used to change or inspect certain TPN attributes, e.g., state (cannot be changed), address attributes, call-back routines, buffer sizes etc. The attributes available will be dependent on implementation. As a minimum it shall be possible to inspect state and to retrieve relevant address attributes.

The programmer must take care to check that attribute changes does not cause unexpected side effects, e.g., when a call-back routine is changed.

Pre-condition: TPN is existing.

Post-condition: Possible change in attributes.

Input parameters: Read/write flag, attribute code, optionally new value.

Output parameters: Old value.

Call-back: None.

Errors: Illegal operation, attribute or value.

5.3.3 Connection point management

5.3.3.1 Create connection point (TLICreateCP)

This service causes a CP description to be created within the TLI. This service creates a general CP object instance that the rest of the A-profile can operate on.

Pre-condition: TLI initiated and TP-network defined.

Post-condition: CP descriptor created in TP_DEF state.

Input parameters:

- TP network.
- Service class requested
- Type of port (listening, connecting)
- Remote NNN if applicable.
- Service instance code
- CP and communication (read and/or write) call-back routines.

Additional, possibly required parameters are:

- Maximum message size if applicable (if not determined by T-profile)
- Buffer parameters if applicable (maximum size etc.)

Input parameters must be specified as appropriate to the service class that is requested and the type of TP network.

Output parameters: CP reference.

Call-back: Communication and connection notifications when connection state changes.

Only state changes not directly a result of another TLI service call and which is defined in the state diagrams shall be reported. Error messages giving causes for unexpected close, e.g., due to jabbering from remote side, shall be reported to the TP network call-back and not to the port call-back routine. If the error results in a TP-network close, it is not necessary to inform the CP call-back about the close.

Errors: Bad input parameters, impossible to create descriptor.

5.3.3.2 Change or inspect CP attributes (TLICPAttribute)

This service can be used to change or inspect certain CP attributes, e.g., state (cannot be changed), address attributes, call-back routines, buffer sizes etc. The attributes available will be dependent on implementation. As a minimum it shall be possible to change call-back routines, check sizes of input and output buffers and to inspect state.

The programmer must take care to check that attribute changes does not cause unexpected side effects, e.g., when a call-back routine is changed.

Pre-condition: CP is existing.

Post-condition: Possible change in attributes.

Input parameters: Read/write flag, attribute code, optionally new value.

Output parameters: Old value.

Call-back: None.

Errors: Illegal operation, attribute or value.

5.3.3.3 Request port connection (TLICPConnect)

This service shall try to establish a connection to or establish as listening, a CP that is already defined.

Pre-condition: CP in state TP_DEF.

Post-condition: CP in state TP_CON, TP_REQ or TP_DEF (if error occurred).

Input parameters: CP descriptor, time-out for attempt.

Output parameters: New state.

Call-back: related to state changes out of the new state.

Errors: Bad descriptor or bad state, attribute errors during connection request (e.g., bad address).

5.3.3.4 Request port close (TLICPClose)

Request the close of a CP in open or connecting state.

Pre-condition: CP in state TP_REQ or TP_CON.

Post-condition: CP in state TP_DEF or TP_CLOSE.

Input parameters: Port descriptor, force flush of output data.

Output parameters: New state.

Call-back: If new state is TP_CLOSE, a call-back will notify of transfer to TP_DEF.

Errors: Bad descriptor or state.

5.3.3.5 Request CP remove (TLICPRemove)

Request the removal of a CP in any state. This implies closing the CP and throwing away buffered data. The CP descriptor will be illegal immediately after a successful call.

Pre-condition: CP existing.

Post-condition: CP removed.

Input parameters: CP descriptor.

Output parameters: none.

Call-back: none.

Errors: Bad descriptor.

5.3.3.6 CP state change call-back (TLICreateCP call back)

This call-back is called when a port changes state independently of an ongoing TLI service call (e.g., open, remote close, error). This will normally be a signal to the application to initiate new operations or new state changes.

Pre-condition: CP existing, state has changed due to external event.

Post-condition: Usually same as state before call, i.e., the TLI shall set the new state before the call-back routine is activated.

Input parameters: CP descriptor, event code.

Output parameters: none.

Call-back: Not applicable.

Errors: Not applicable.

5.3.4 Data management

Data can be transmitted on a CP or received from a CP when it is in the TP_CON state. The following services are used to send or receive data.

5.3.4.1 Buffer handling

The TLI must provide mechanisms for in- and output buffer management. This standard does not specify how this shall be done. In the following it is assumed that a read from port returns a buffer, and that a write to a port consumes a buffer. It is also assumed that each read event delivers one data buffer and that the calling application (A-profile) can handle cases where write calls are not able to output data immediately.

The implementation must consider efficient buffer management (minimise copying), buffer pool management (dynamic or static allocation of buffers) and operating system related issues (e.g., allocation in the write sub-system and de-allocation from protocol level).

In the following service specifications, it is assumed that a read or write buffer follows the relevant events or service calls and that the ownership of this buffer follows the flow of control. It is also assumed that creation of write buffers and deletion of read buffers must be done explicitly.

5.3.4.2 Read from CP (TLICreateCP call back)

This service is implemented as a call to the port call-back routine when one buffer is ready for reading. This applies to both stream and message data. When this event is delivered, the corresponding input buffer should follow the event. If this is the case, the buffer must normally be freed by the application after processing.

Pre-condition: CP open with read data in a buffer.

Post-condition: No change, data buffer follows event.

NOTE – It is assumed that the higher layer application is the new owner of the read data buffer ant that it must explicitly free the buffer, either in the event handler or somewhere else.

Input parameters: CP identifier and read buffer.

Output parameters: Not applicable.

Call-back: Not applicable.

Errors: Not applicable.

5.3.4.3 Release read buffer (TLIDeleteBuffer)

This service deletes a read (or pre-allocated write) buffer that has been owned by the higher layer application.

Pre-condition: Buffer owned by application.

Post-condition: Buffer released to TLI pool.

Input parameters: Buffer.

Output parameters: Not applicable.

Call-back: Not applicable.

Errors: Not owner of buffer.

5.3.4.4 get write buffer (TLIAllocateBuffer)

This service creates a buffer that can be used for formatting of an output message.

Pre-condition: TLI initialised.

Post-condition: Buffer owned by application.

Input parameters: Maximum size of buffer.

Output parameters: Buffer.

Call-back: Not applicable.

Errors: Buffer pool empty.

5.3.4.5 Write data to a CP (TLICPWrite)

This service is used by the application to request the sending of a formatted stream of octets or a message out from a specified CP.

The service will normally queue data for output, but if queues are excessively long and the CP has a low priority, the calling application can be instructed to stop sending data until notified. In this case no part of the buffer will be queued for output. The notification that write can be attempted again is through a call-back from which this service can be invoked again. This is illustrated in the below figure, which represents sub-states of port state TP_CON.

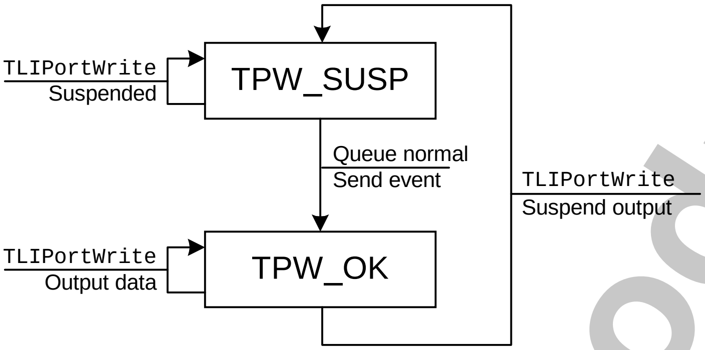


Figure 7 - Write sub-states

Pre-condition: CP is open and ready to accept write data. A buffer has been formatted with output data.

Post-condition: CP has data queued for output or port in write suspended sub-state. In the former case, the buffer has been consumed.

Input parameters: Buffer (including length of data), port.

Output parameters: Operation status (full - wait for call-back or ready for more).

Call-back: If state is full, call-back when CP state is ready.

Errors: Bad CP or buffer.

5.3.4.6 CP write call-back (TLICreateCP call-back)

This call-back is used to signal that the application can write more data to the port.

Pre-condition: CP output queue ready.

Post-condition: Buffer consumed.

Input parameters: CP identifier, event code.

Output parameters: Not applicable.

Call-back: Not applicable.

Errors: Not applicable.

6 T-profile for redundant Internet

6.1 Introduction

6.1.1 Rationale for the use of Internet protocols

IP, currently in version 4, provides message and stream transmission services over any kind of LAN, MAN or WAN networks. The IP protocol is independent of the underlying network technology such as Ethernet, FDDI or ATM, thus enabling interoperability among these different network technologies and providing a safe migration path for the use of future technology.

The IPV4 transport profile specified in this clause relies on the IP network protocol and will also be independent of the underlying network technology.

This approach allows flexibility towards the choice of the network technology and ensures longevity of the system level network implementation due to, e.g. evolution of the network technology. The transport profile implementation will also take advantage of the many different commercial implementations that already are available on the market. This approach also facilitates the use of widespread Internet applications such as network management, time distribution and terminal emulation. In addition, the transport profile will benefit from the new emerging mechanisms and services that have been in development for several years by the Internet Engineering Task Force (IETF), e.g., IP version 6 (IPV6).

6.1.2 Relationship to basic Internet protocols

The T-profile is based on the Internet transport profile that provides:

- A connection oriented stream transfer protocol: The Transmission Control Protocol (TCP),

  A connection-less mode message transfer protocol: The User Datagram Protocol (UDP).

TCP supplies a robust service of transport including the following major properties:

- end to end transfer,
- full duplex data exchange,
- error detection and retransmission,
- guarantee of in-sequence delivery without duplication,
- flow control,
- congestion control.

TCP supports only a unicast transfer mode.

UDP provides a best effort strategy for message delivery (without guarantee). UDP supports either a unicast, multicast or broadcast transfer mode.

Both TCP and UDP protocols support client-server type communication.

Internet transport protocols satisfy a subset of the required communication services. Unreliable short message transfer can be achieved through UDP protocol while reliable stream transfer can achieved through TCP protocol. Reliable message transfer can be achieved through an additional protocol built over TCP. Redundancy can be achieved with the use of two independent networks (transmission channels). Additional protocols are also required for this. Neither TCP nor UDP provides priority. A simple priority mechanism will be provided through the use of the TCP urgent mechanism.

6.1.3 Rationale for use of Ethernet

Ethernet has been defined as the physical and data-link layer for this standard due to its wide use in the industry, its availability for various hardware platforms and its relatively high robustness. It has also been a major factor that Ethernet is a typical mainstream technology that is continuously being developed and enhanced.

Ethernet is not able to provide quality of service control and it uses a stochastic network arbitration scheme. These factors mean that it cannot be used for hard real-time control where accurate prediction of network latency is required. In general, the latency and performance of the Ethernet based network will depend on the load condition of the network.

For system integration purposes, where required latency is above 100 milliseconds and no hard real-time deadlines are needed, Ethernet should perform well under most reasonable load conditions. However, care must be taken so that load is not excessive in any operational state.

When switched Ethernet is used, many of the above assumptions become void. Switched Ethernet provides a full cross-bar switch between all host computers connected to a switch and will not normally use the stochastic medium access mechanism. This means a much higher degree of determinism in the network. However, due to the relatively high complexity of modern workstations, the system integrator should still not rely on deterministic end to end delay for this protocol. Unless special precautions are taken, the limit of 100 ms as the shortest expected response time should be observed.

6.2 Transport profile architecture

6.2.1 General

The transport communication, time and management services will be provided through the use of a set of protocols, based on Internet protocols, located either at the OSI communication layers (layers 3 to 4) or at upper layers (layers 5 to 7).

The general PISCES protocol architecture at the system level network, using the T-profile defined in this part of the Standard, is illustrated in the following figure. The shaded A-profile shows that it is not covered in this part of the Standard but is still part of the complete protocol architecture.

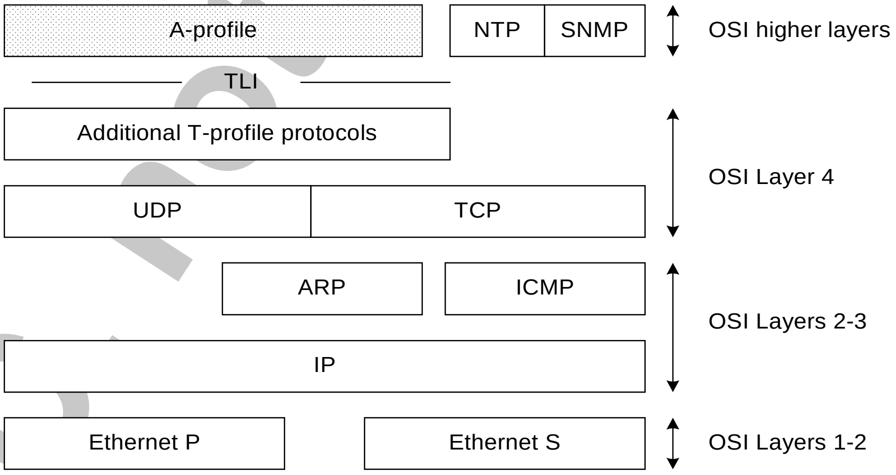


Figure 8: General transport profile architecture

This architecture satisfies the needs currently identified at the system level network aboard a ship and also makes allowances for the future needs associated with enhanced multimedia applications.

6.2.2 Redundancy

Redundancy is implemented by using two physical network interfaces and two independent Ethernet (Primary and Secondary). Each of these networks has different Internet network addresses. However, only one Internet protocol stack is normally used in the host computer. This is made possible by using the multi-homing facilities of the Internet protocol stacks.

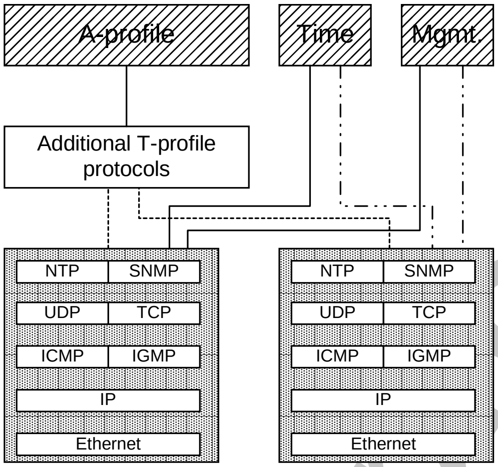


Figure 9 - Redundancy implementation

Higher level protocols (NTP and SNMP) will operate independently on the two networks with no modification of basic protocol or services. This means that the time and network management services will need to be aware of and interface to both networks, i.e., regard each host computer as having two distinct network addresses, one on each network.

The A-profile will make use of additional T-profile functionality as described in this clause. This means that the A-profile (and, thus, the LNA and MAU) will relate to only one logical network (using one TP network address and one network node number for each host computer).

The additional T-profile services will cover both UDP and TCP. The basic principle is that all data is sent on both physical networks simultaneously and that duplicate data is filtered out at the receiving side. This provides zero switch-over time between the two networks and continuous availability.

Multicast/broadcast services will use UDP broadcasts. Each physical network must be set up so that broadcasts can reach all nodes on the network, e.g. any switches and routers should normally not do broadcast filtering. To handle duplicate messages, a sequence number will be added to all UDP messages so that the last arriving of the duplicate messages can be removed. This will also ensure that all messages arrive in sequence.

Reliable stream services will use an octet count to do duplicate removal. Reliable message based traffic uses the same principle and message assembly is done only after the duplicate octets have been removed. After errors on one network, where redundant communication is resumed at a later stage, the T-profile uses a special protocol to synchronise octet streams after reconnect.

6.2.3 Communication services

6.2.3.1 TP network types

The below table lists the TP networks that are defined in this clause. These TP networks are identical except that the first group has predefined IP network addresses. The second form allows user specified network addresses. Note that the second form shall be able to handle all classes of IP networks while the first is defined as a class B network.

The TP networks are divided into MAU-MAU (stream) and LNA-LNA (message) type networks. For the first group it is necessary to specify TCP port as part of the address parameters. For the second group, the TCP/UDP port is implicit in the selected service.

Table 8 - Redundant TP networks

Code	TP network address format	NNN address format	Notes
LNADR_IPV4R (129)	IPV4 standard redundant address for LNA. The address field is 0 octets long, being void.	The host Internet address with the network part, all zeros.	The two networks are the ones defined in clause **Error! Reference source not found.**.
MMADR_IPV4R (131)	IPV4 standard redundant address for MAU. The address field is 0 octets long, being void.		The two networks are the ones defined in clause **Error! Reference source not found.**.
LNADR_IPV4RE (130)	IPV4 free redundant address for LNA. The address field is 8 octets long, giving the Internet network addresses in binary format, primary network first, then secondary. The node number part of the addresses shall be all zero bits.		The two networks (of the same class) shall be as defined in clause **Error! Reference source not found.**.
MMADR_IPV4RE (132)	IPV4 free redundant address for MAU. The address field is 8 octets long, giving the Internet network addresses in binary format, primary network first, then secondary. The node number part of the addresses shall be all zero bits.		The two networks (of the same class) shall be as defined in clause **Error! Reference source not found.**.

6.2.3.2 Services supported

The services defined in clause 4.3.3 shall be implemented, with the exception of the MAU-LNA message service. The services may, however, be implemented in a number of different modules where suitable service sub-sets can be linked into the application, either a MAU or an LNA. This may mean that several different TP-network instances use the same network code.

6.3 Message formats

6.3.1 General format

The A-profile (part 401 of this standard) defines a general message format that shall be used by all the A-profile protocols. This message format will also be used by this T-profile. The general message header is shown in Table 9. Refer to IEC 61162-401 for a detailed discussion of the different components of the message.

Table 9 - General message format

Octet #	Data type	Field description
0 – 1	`word16_m`	`MV4_MAGIC`
2 – 3	`word16_m`	Message type
3 – 4	`word16_m`	Priority
4 – 5	`word16_m`	Total length of message in octets (x+1)

Octet #	Data type	Field description
- x		Data

6.3.2 Special reliable link messages

To synchronise the two channels of a redundant link and to inhibit the creation of more than one communication channel in a peer to peer link, a special message is sent ahead of all other traffic. This message is described in Table 10.

Table 10 – Special reliable hello message format

Octet #	Data type	Field description
0 – 1	word16_m	MV4_MAGIC
2 – 3	word16_m	LL_HELLO (0x0001)
3 – 4	word16_m	Priority
4 – 5	word16_m	Total length of message in octets (16)
6 – 7	word16_m	Type of network
8 - 11	word32_m	NNN of sending host computer
12 – 15	word32_m	Next octet number

The message identity code has the value one. The range from 1 and up to 255 is reserved for management messages.

The priority is not used and can be discarded. The total length is fixed at 16 octets.

The type of network is a code specifying the TP-network used by the sending host computer (it will typically be one of the codes of Table 8). It can be used to detect non-redundant hosts on a redundant network.

The NNN is the network node number of the transmitting host computer. It is used to break ties if two host computers simultaneously act as client and server during a peer to peer connection establishment.

The next octet number is used to synchronise octet streams when a redundant link is re-established after a break in one of the communication channels. It is not used for non-redundant networks. It shall be zero for the first connection attempt. When a broken channel is reconnected it shall contain the sequence number of the data octet immediately following the LL_HELLO message. The first data octet transmitted on a channel is counted as zero (consistent with the zero value of the octet number in the initial message).

6.3.3 Broadcast messages

To filter out duplicate messages, the NNN together with the sequence number is added to the broadcast messages by the T-profile. This is done as a prefix to the message header shown in Table 9. Thus, the complete message format for broadcasts is as shown in Table 11.

Table 11 - Broadcast message format

Octet #	Data type	Field description
0 – 3	word32_m	NNN of sending host computer
4 – 5	word16_m	Sequence number for this broadcast port

6 – 7	word16_m	MV4_MAGIC
8 – 9	word16_m	Message type
10 – 11	word16_m	Priority
12 – 13	word16_m	Total length of message in octets (x+1)
- x		Data

The NNN is the network node number of the sending host. It is used to differentiate between host computers. The sequence number is used to identify messages from one host computer being sent on two redundant channels.

The T-profile shall strip the message of the leading NNN and sequence number before passing it up to the higher layers. The T-profile shall also check message consistency (magic number and message length).

6.4 Reliable stream protocol

6.4.1 Introduction

The reliable stream protocol is also the basis for the reliable message protocol. It uses a TCP/IP connection between the two communicating host computers and provides a reliable, sequential and bi-directional octet stream.

6.4.2 Connection management

6.4.2.1 Difference between peer-to-peer and client-server

The reliable stream protocol shall support two connection modes: Peer-to-peer and client to server. Peer-to-peer communication is established between a client (connecting) and a server (listening), but both parties may simultaneously act both as client and server. In the latter case the two established connections shall be merged into one by the T-profile. In the client-server model the two links shall remain distinct.

As client-server connection establishment can in this respect be looked at as a sub-set of the peer-to-peer establishment. This T-profile also implements the two services this way: A CP flagged as a peer to peer type CP will be subjected to additional processing to make sure that only one communication link is established between the parties. This processing is not used on a client-server type CP.

The next clause gives details on the establishment of peer to peer connections. The client-server establishment use normal TCP/IP mechanisms with a listening and a connecting CP. Redundancy is handled by using the same mechanisms as the peer-to-peer connection (6.4.2.3 and 6.4.2.4).

6.4.2.2 Resolving peer-to-peer connection conflicts

Each of the two channels shall be established independently and concurrently. The state changes for each channel are described in the below diagram.

The left side shows the state changes for the client side, while the right side shows the state changes for the server side. Prior to starting the client process, the client shall check if the link is already open. The identity (MyId for local host computer or Id for remote host computer) is the network node number. It is guaranteed to be unique and a simple check on relative numeric value can be used to break a tie when both sides try to initiate a connection. The result shall be that the side with the highest NNN shall succeed in being the client. The LL_HELLO message is used to send the identity codes.

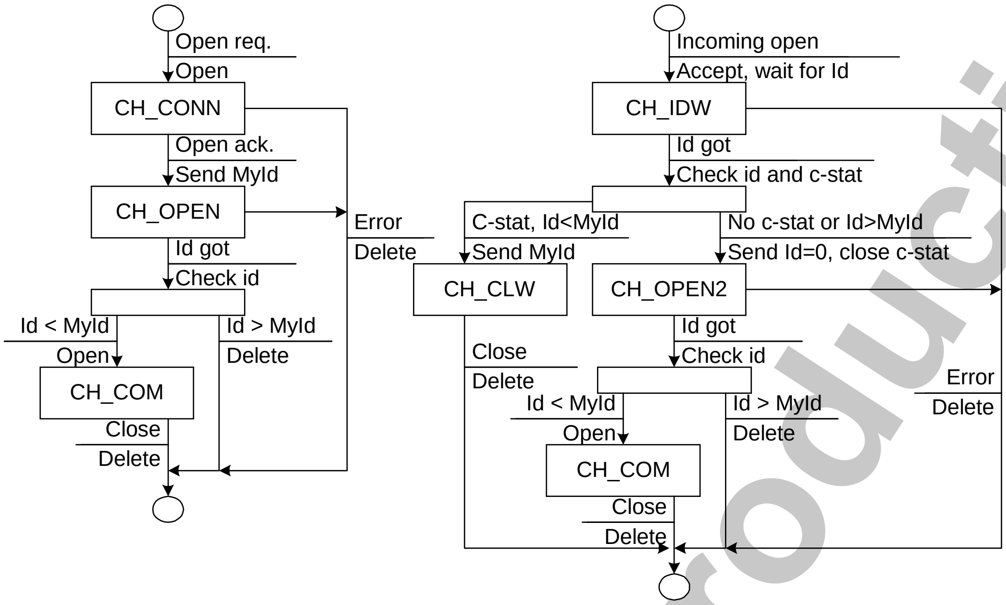


Figure 10 - Connection establishment

6.4.2.3 Redundancy modes

The T-profile shall provide graceful degradation, i.e., it shall maintain communication even if one of the redundant channels goes down. The following state diagram shows the possible connection states.

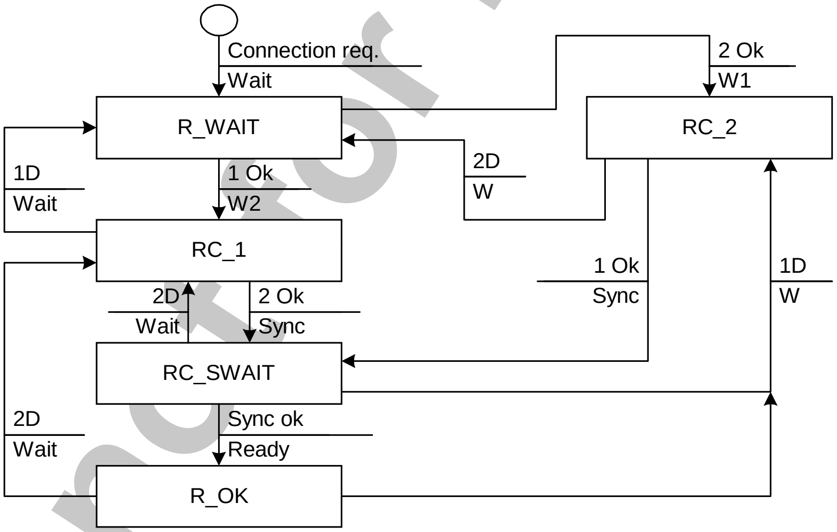


Figure 11 - Redundancy communication modes

The events are channel down (e.g., 1D) or channel just connected to (e.g., 1 Ok).

The states are:

- R_WAIT: No communication established. Waiting for connection on both channels.
- RC_1 and RC_2: Connection established on primary or secondary channel respectively. Data transfer enabled on the connected channel.
- RC_SWAIT: Both channels connected to. Data reception enabled on the channel corresponding to the state just left (RC_1 or RC_2). Data is transmitted on both channels.

Waiting for data synchronisation on the other channel before full redundant reception is enabled.

- R_OK: Full redundant communication established.

The synchronisation wait (RC_SWAIT) is necessary to check that the remote host computer actually is transmitting correct information on both links before the second link is accepted as a full replacement for the first established link.

In all communication modes, data is transferred to the higher levels as soon as they have been received. The T-profile shall not delay data.

6.4.2.4 Channel synchronisation requirements

The T-profile shall provide a mechanism for consistency checks between the two channels. Different mechanisms can be employed, but the mechanism shall implement a check to find (with a suitable probability) the following errors in one of the channels:

a) It shall be able to detect the omission of a data block.

b) It shall be able to detect the repetition of a data block.

c) It shall be able to detect the wrong ordering of two data blocks.

d) It shall be able to detect the insertion of a number of random octets, e.g., some number of zero value octets.

These errors are the most probable that can occur in the sending T-profile. It is not necessary to detect errors due to transmission problems on the physical network. This is reasonably well handled in the TCP/IP and Ethernet layers.

NOTE 1 - Requirement 3 is probably the most difficult to handle with a reasonably effective check. It requires relatively fine grained check-point intervals or some form of positional memory (e.g. a CRC) in the algorithm.

Any errors detected by the synchronisation algorithm shall immediately cause the communication link (both channels) to be closed and an exception to be generated for the higher layers. It is the responsibility of the higher layers to reconnect the link.

NOTE 2 - The link shall also be closed when an error occurs during reopening of a redundant link in the RC_SWAIT state. It is generally not possible to prescribe a safe action in this case, even when one channel is operating.

The LL_HELLO message contains the necessary information for restoration of a broken redundancy link. It contains the octet sequence number for the immediately following data.

Annex B provides an example of a possible implementation of such an algorithm.

6.4.2.5 Priority

TCP/IP can to a certain degree support two priority levels: Normal and urgent. Urgent priority is achieved by setting the out of band (urgent) flag after having transmitted a block of urgent data. Normal priority data shall not use the out of band flag.

NOTE - The actual effect of using out of band flags to implement priority will depend on the implementation. The TCP/IP stack is required to make a best effort to send this data before other data, but implementations vary in how they achieve this.

To make this mechanism work, only urgent messages shall be sent on an urgent link and only normal priority messages on a normal priority link. This means that a link that shall support both normal and urgent priority data (e.g., the LNA-LNA message link) must consist of four TCP/IP connections: Two for redundancy in each of the two priority links. Each pair of connections (normal or urgent priority) can be handled as one link on which the T-profile sends a certain sub-set of data from the higher level CP.

Data links that shall support only one priority level (urgent or normal, e.g., for stream links) should normally not create the other connection. The parameters determining the priorities to be supported are input to the TLI.

Low priority cannot be supported by TCP/IP as only two priority levels exist.

6.4.3 Data transmission

6.4.3.1 General

Data is transmitted on the active links as soon as they are open. This means that data transmission shall start in RC_1 and RC_2 modes and that a synchronisation is necessary when fully redundant transmission starts at some later stage.

The T-profile must keep track of the octet count used for transmission so that the correct LL_HELLO message can be sent when a link opens. It must also make sure that the transmitted octet count is consistent with the actual octets transmitted.

The T-profile shall generally send data as soon as data is requested sent by the higher layers. The T-profile may, however, buffer data internally or refuse transmission if the data channel is congested.

6.4.3.2 Transmission of urgent data

The urgent flag shall be set after each complete block of data handed over to the TCP layer, e.g., after each write call.

6.4.3.3 Graceful degradation between priority levels

If one of the links (urgent or normal) fail, transmission can continue on the other link provided that the following requirements are fulfilled:

- Normal data shall not be transmitted on the urgent link (signal normal link failure to higher layers).
- Urgent data can be transmitted on the normal link, but no normal messages can be transmitted on this link when urgent messages are transmitted (signal normal link failure to higher layers).
- Connection reestablishment shall be attempted immediately.
- As soon as both links are re-established, the urgent messages shall be redirected to the urgent link and normal link functionality resumed.

Graceful degradation is only provided between urgent and normal links established at the same time. The T-profile shall not reallocate link descriptors, e.g., to give a low priority stream to a normal priority message link. This shall be handled, if necessary, by the higher layers.

6.4.3.4 Local link close

The link close command from the higher layer shall cause all data still being in the communication pipeline to be transmitted before the link is physically closed, unless a throw flag has been included in the close command. In the latter case, the link shall be closed immediately. Both redundant channels shall be closed at the same time. This may mean that the link cannot be reopened immediately after a close (if buffers need to empty). The throw flag can be used to enable immediate reopening.

NOTE – The normal TCP/IP close will normally allow buffers to empty before the close is completed. The throw flag requires special handling. This has to be checked during implementation. Note also that TCP/IP sockets may linger and that this may have to be handled in the implementation by setting, e.g., no-linger flags.

6.4.3.5 Remote link close

The same provisions as for the local close also applies to remote link close. The CP may or may not have to flush its output buffers. Note in particular that data may be arriving although the connection do not allow the transmission of data (to empty the sender buffers).

NOTE – Errors in the communication layer may be detected only during write operations for some error modes. Note also that the keep-alive timer used in TCP/IP is not very useful for control systems. the time-out is typically several minutes or sometimes hours.

6.4.4 Data reception

6.4.4.1 General

Data received shall be transferred to the higher layers as soon as they have been received on one of the open channels. The synchronisation mechanism described in 6.4.2.4 shall be used to check consistency between channels, but the mechanism shall not delay the transfer of data to higher layers.

6.4.4.2 Flow control

The receiver may employ normal TCP/IP flow control mechanisms if a link becomes congested, i.e., stop doing read on the relevant connection point. This is, however, not legal on urgent links. These shall be processed immediately and any congestion control must be performed on other link types.

6.4.4.3 Priority

Urgent data received on an urgent link must process the out of band flag. This means that, e.g., urgent signals from the operating system must be trapped. It also means that all data on an urgent link must be read, up to the last pending urgent flag.

6.4.4.4 Graceful degradation between priority levels

The receiving end should normally only get urgent messages on the urgent link. However, observe the possible change from normal to urgent in the case of an urgent link failure.

6.4.5 Link close

The provisions described in 6.4.3.4 apply also on the listening side.

6.5 Reliable message service

6.5.1 Introduction

The reliable message service shall use the reliable stream protocol and use the general message format (clause 6.3.1) to find message boundaries. This means that most services described in the previous clause are implemented in the same manner.

The exceptions are noted in the following clause(s).

6.5.2 Data reception

The T-profile is required to only deliver complete messages, one and one at a time, to the higher layers. This may require buffering of data to make sure that a complete message has been assembled.

6.5.3 Data transmission on the urgent link

The T-profile shall set the urgent flag only on message boundaries. If, for some reason, two or more messages are transmitted immediately after each other, the transmitter may set the urgent flag only after the last message.

6.6 Unreliable message service

6.6.1 Introduction

The unreliable message service uses the UDP broadcast mechanism. This is a connection-less and unreliable service. However, the TLI shall be defined in such a way that the users of this service can use the normal concepts of connection points and data input and output on connection points.

6.6.2 Connection management

6.6.2.1 Introduction

There is no protocol level connection management associated with the unreliable message service other than maintaining and checking sequence number for all messages. However, the TLI specifies that a CP shall be associated with each listening or transmitting broadcast address.

This standard specifies five different listening UDP ports for broadcast messages. Each of these ports can be associated with exactly one transmitting and one receiving CP.

6.6.2.2 Listening CP

The T-profile shall maintain state associated with each remote transmitting host computer that transmits to a specific UDP port. The state is limited to the identity (NNN) of the remote host computer and its current sequence number. When opening a new listening CP, the T-profile shall reset this state information to no remote host computers. As soon as a message from a previously unlisted host computer arrives, the state shall be updated with the identity of that computer and the sequence number of the message.

A remote transmitting host computer may go off and come on line due to a reset or other internal exception. This will cause a reset of the sequence number to one. The receiving end must be able to detect this and continue reporting incoming messages from the new sequence number. In systems that employ routers and alternative transmission paths, one may see duplicates of UDP messages, and hence, out of sequence messages, without an exception on the transmitting end. This situation is extremely unlikely for a LNA type ship control system and this standard prescribes the following handling of out of sequence packets:

a) Out of sequence messages on the same physical network shall normally be regarded as the result of a transmitter exception and lead to a reset of the listener's sequence number counter.

b) Exceptionally, if the newly arriving message has a high sequence number (significantly higher than zero) which is close to the current sequence number, the message may be thrown away.

NOTE – "high" in this context should be at least 100. "Close" should be less than 10. There are no firm limits here and the decision is left to the implementation. The duplication situation is much more unlikely than a host computer reset and the latter situation must be given priority.

The standard defines the multi-cast service as unreliable and potentially with messages out of sequence, so the A-profile is constructed with these limitations in mind. This means that the alternative two above should be used with great care.

This standard specifies five different listening UDP ports for broadcast messages. Separate state shall be maintained for each port, i.e., distinct sequence number series will be used for each port.

Only one CP is allowed for each UDP port. The T-profile need not support duplication of incoming messages to possibly several ports.

The sequence number for the transmitting host shall be the same for both links, when redundant communication paths are used.

6.6.2.3 Transmitting CP

The T-profile shall maintain a sequence number for all outgoing messages on a given UDP port. The sequence numbers shall be separately maintained for each port, but it shall be common to redundant transmission paths for that port. It shall be set to one for all ports when the TP-network is established. Closing and opening a CP on a specific port shall not reset the sequence number for that port.

The sequence number shall be incremented with one for each new transmitted message.

6.6.2.4 Redundancy management

This service shall transmit and listen on both redundant networks unless instructed by higher level management services to use only one of the networks. There are no particular connection or redundancy management provisions required for this service.

Although sequence number management shall be distinguished between on UDP ports, but it shall be common to redundant links. More details are given in the following clauses.

6.6.3 Message length limitation

Outgoing messages are limited in size to that supported by the low level network. For IP over Ethernet, the maximum transmission unit (MTU) is 1500 octets [RFC 894]. The limit is reduced by various headers as listed in the table below.

Table 12 - Maximum message size

Protocol header	Size (Octets)
IP header	20
UDP header	8
T-profile broadcast header	6
Sum	34

The maximum payload for higher layer protocols is 1466 octets.

6.6.4 Data transmission

Outgoing messages shall be checked for maximum length before they are transmitted. They shall be formatted with an additional header containing the NNN of the transmitting host and a sequence code as described in 6.3.3. The first message on a newly established transmitting CP shall use the sequence code one.

NOTE - It is suggested that the TLI uses some form of data buffer management that reduces the need to copy data back and forth to insert the necessary header information.

The formatted message shall be transmitted on both the primary and the secondary network in two identical copies.

The transmitting service is not required to retransmit messages when errors occur locally (e.g., output buffers full). However, the T-profile shall make a best effort to transmit all outgoing data (e.g., by allocating sufficient memory for buffers).

6.6.5 Data reception

Messages received on a listening and redundant CP shall be filtered so that:

- The minimum of duplicate messages is delivered to higher levels.
- The minimum of out of sequence messages is delivered to higher levels.
- Messages are delivered with minimum delay in the T-profile.
- The receiving part of the T-profile shall cater to the situation where a remote host is reset so that sequence numbers are reset (6.6.2.2).

The filtering shall be done based on the NNN of the transmitting host computer and the sequence number. The T-profile shall immediately deliver a message to the higher layer when a message with a higher sequence number is received on one of the two networks, even when this number shows that one or more messages can have been lost. Any lost messages shall be handled by the higher layers.

NOTE 1 – The reset of transmitting host causes problems which are “contrary” to the filtering requirements. Priority must be given to the reset issue as that is more likely in most contexts. However, one scan examine the sequence numbers on the same network together with those on the redundant network to be able to make a better decision on this issue.

NOTE 2 - The sequence number will start at one (see previous clause). This means that the sequence number of a newly established listening CP can be set to zero to enable the first message to pass through. Note also that sequence numbers are 16 bit and that a wrap around will cause the number zero to be used. The wrap around will also have to be handled in conjunction with message filtering.

NOTE 3 - As two independent networks are used, there is theoretically a possibility that out of sequence messages can be received, even on a local area network. This can occur if one or more messages are lost on the fastest transmitting network. It is recommended that this possibility is ignored by the T-profile and that it passes sequential messages (with possible gaps) to the higher layers as soon as possible.

6.7 Time services

6.7.1 Introduction

This standard uses the Network Time Protocol – NTP [RFC 1305] and the Simple Network Time Protocol – SNTP [RFC 2030] to implement a distributed time reference in the IEC 61162-4 network. A short overview of these protocols can be found in Annex C.

6.7.2 Time Management Architecture

This standard requires that each of the two redundant networks shall have at least one time server, each supporting NTP (version 3 or higher). The time servers (program modules) shall reside in host computers that have a connection to a reliable UTC source, e.g., a GPS receiver.

The two time servers should use independent time sources.

The time servers shall send all multicast messages on both networks and shall also be able to serve unicast or anycast queries arriving at any of the two networks. The time servers shall support all SNTP services used by clients (unicast, multicast and anycast). The two networks shall, otherwise, be regarded as two separate networks with different IP address ranges.

The two time servers should work together in symmetric peer mode (NTP) to back each other up in case the time source for either fails.

NOTE 1 - Bringing up a NTP primary server requires a radio or satellite receiver or modem. Support for the commonly available GPS receivers is included in the default configuration of many of the commonly available NTP daemons, e.g., `ntpd` for UNIX. Refer to the manuals for the various NTP modules for more details on configuration.

NOTE 2 – A GPS receiver requires support in the form of a serial port. On a typical work-station the timing jitters contributed by the serial port hardware and software driver can accumulate to several milliseconds. A set of special line disciplines and stream modules can be configured in the operating system kernels in order to reduce these errors. These routines intercept special characters or signals provided by the radio or modem clock and save a local timestamp for later processing.

6.7.3 Client interfaces

Clients shall have an interface to both time servers. The client shall select the service level that gives the time accuracy as required by the server. SNTP is normally sufficient for all but the most strict requirements. It is recommended that the client uses the SNTP anycast or multicast mode to avoid unnecessary configuration in the system.

The expected accuracy for the various service levels shall be specified by the system integrator. This will dependent on accuracy of time sources, of configuration files and of general network configuration.

6.7.4 System integration requirements

The system integrator shall determine the required time accuracy and construct the servers and the architecture so that this accuracy requirement is met. It shall also be specified what is the accuracy and what service the client can use to achieve the accuracy.

6.8 Network management services

6.8.1 Introduction

This T-profile will specify physical network management over the SNMP protocol [RFC 1157]. An informative overview of this protocol is included in Annex D. System management shall be performed by the use of the Management Information Base described in [RFC 1213].

6.8.2 Physical architecture

The actual physical architecture shall be determined by the system integrator based on the requirements set by authorities, class and others at the time of system design. However, unless such explicit guidelines have been developed, the following requirements apply.

Two network management agents (NMA) shall be deployed, one on each of the two redundant networks. Each NMA shall keep a complete overview of one of the networks. Each NMA may also connect to both networks to achieve a higher degree of robustness.

Each client shall be able to report MIB information to both NMA.

6.8.3 Required safety precautions

The MIB supported by any NE must not allow unauthorised personnel access to writeable parameters that can endanger system safety (e.g., routing or address parameters). This safety feature may have to be implemented in NE as well as in NMA.

Normal NE need not generally use encryption or special authentication procedures as they reside on a protected network.

A firewall or any NE that resides on the IEC 61162-4 network as well as any other network where unauthorised personnel can access the MIB must be protected from unauthorised tampering via the second network.

6.8.4 NMA requirements

Each NMA shall be able to detect and report anomalies that are indicative of developing or imminent problems in a way that is useful to the operator of the NMA. Where appropriate, the NMA should be connected to a general alarm and monitoring system.

6.8.5 Network element requirements

The MIB that shall be implemented by the client is defined in [RFC 1213]. In particular, the system, interfaces, icmp, tcp and udp groups are mandatory for all equipment that support these protocols, i.e., host computers in the IEC 61162-4 network.

The system integrator shall ensure that other network elements, e.g., routers and switches, provide the relevant MIB. These are usually the system and interfaces groups.

7 T-profile for non-redundant Internet

7.1 Introduction

This clause defines a T-profile intended used over non-redundant local area networks. The T-profile is mainly intended for devices that do not require redundancy, but need to be connected into a ship control network anyway, e.g., various data acquisition devices. This T-profile is interoperable with the redundant T-profile described in the previous clause.

The following table lists the TP network codes used for this T-profile.

Table 13 - Non-redundant Internet TP networks

Code	TP network address format	NNN address format	Notes
LNADR_IPV4 (120)	IPV4 standard non-redundant address for LNA. The address field is void.	The common host computer part of the addresses. Network part- all zeros.	The host computer can only be attached to the primary channel of a redundant network as defined in clause **Error! Reference source not found.**.
MMADR_IPV4 (122)	IPV4 standard non-redundant address for MAU. The address field is 2 octets long, containing the port as a word16_m.		The host computer can only be attached to the primary channel of a redundant network as defined in clause **Error! Reference source not found.**.
LNADR_IPV4 (121)	IPV4 free non-redundant address for LNA. The address field is 4 octets long, giving the Internet network addresses in word32_m format. The node number part of the addresses shall be all zero bits.		The network shall be as defined in clause **Error! Reference source not found.**.
MMADR_IPV4 E (123)	IPV4 free non-redundant address for MAU. The address field is 6 octets long, giving the Internet network addresses in word32_m format and then the port number in word16_m format. The node number part of the addresses shall be zero.		The network shall be as defined in clause **Error! Reference source not found.**.

7.2 Relationship to redundant Internet protocol

This T-profile shall support the same services as the redundant network protocol as far as they can be mapped to a non-redundant service. All provisions of clause 6 shall apply also to this T-profile with the exceptions as described in the following clauses.

7.3 Redundancy

Redundancy requirements as described in 6.2.2 do not apply to this T-profile. The host computer can be attached to either the primary or the secondary network. However, the host computers specified TP-network must match that of the network it is attached to.

The time and network management modules of the T-profile need only consider the one network the host computer is attached to.

The TLI shall make the differences between the non-redundant and the redundant network invisible to the higher layer protocols.

7.4 Reliable stream protocol

All provisions of 6.4 applies, with the following exceptions:

- Redundancy as described in 6.4.2.3 cannot be supported. The connection is in the R_WAIT or RC_1 modes. RC_1 is used for both primary and secondary network attachment. The T-profile shall send the `LL_HELLO` message as prescribed.
- Channel synchronisation as described in 6.4.2.4 does not apply. This is consistent with the CP being only in the RC_1 state. Data is sent to higher layer without consistency checks. However, the T-profile shall send the initial `LL_HELLO` message.
- Data is sent and received only on one channel.

7.5 Reliable message service

All provisions of 6.5 applies.

NOTE - The implementation of the reliable stream service shall hide the existence of redundancy from this layer of services.

7.6 Unreliable message service

All provisions of 6.6 applies, with the following exceptions:

- Messages are sent and received on one network only.

7.7 Time services

All provisions of 6.7 applies, except that messages are sent and received on only one network and that only one time server is normally necessary.

7.8 Network management services

All provisions of 6.8 applies, except that messages are sent and received on only one network and that only one NMA is normally necessary.

8 T-profile for wide area networks (MAU-LNA communication link)

8.1 Introduction

To facilitate remote integration tests, this clause describes how a MAU on a host computer located anywhere on the world wide Internet can be connected to a central LNA used for testing an integration of a control system.

The WAN protocol can also be used to integrate very simple MAUs (with limited communication facilities) to a system level network by using a simple point-to-point TCP/IP link between the MAU and an LNA on a larger computer.

8.1.1 General architecture

This clause assumes that one or more LNAs are set up on a local area network with the possibility for the host computers on that LAN to accept incoming TCP/IP connections from anywhere in the world. This is illustrated in the below figure.

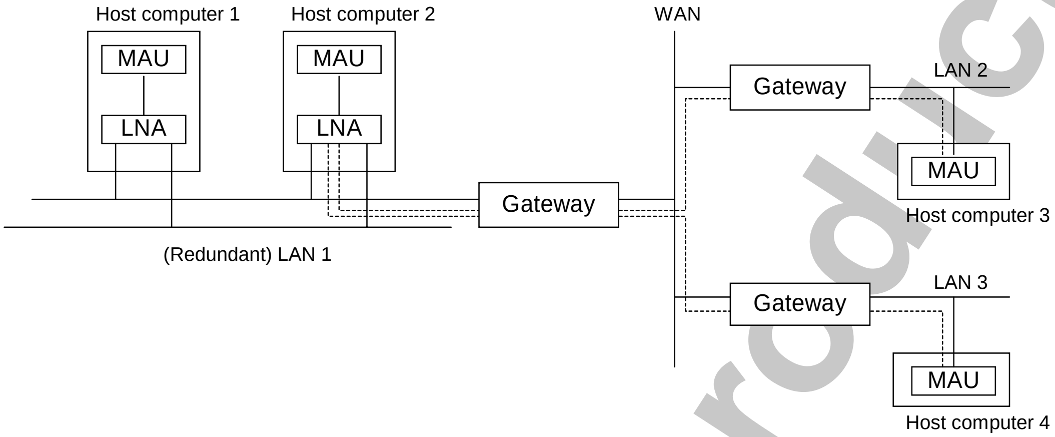


Figure 12 - WAN Architecture

From the MAU's and the LNA's viewpoints, the architecture will look as if all the MAUs were located on the same control network.

NOTE - It would be possible to implement a WAN type LNA network with the use of, e.g., Internet Group Management Protocol (IGMP) and multicasts, but this requires a more complex set-up.

8.1.2 TP-network and NNN code definitions

The MAPI and the LNA must be able to use the TCP/IP MAU-LNA link with the same TLI as is normally used between MAU and LNA (LNAC variant of TLI). The following TP networks are defined to handle this communication.

Table 14 - WAN TP networks

Code	TP network address format	NNN address format	Notes
MAUADR_IPC (1)	Not used (zero).	Application (MAU) specific code (e.g., process identity).	Not defined in this standard. Reserved for LNAC type TLI.
MAUADR_TCP (2)	Not used (zero)	The full Internet address of the client host computer.	Defined in this clause.

8.2 Relationship to non-redundant Internet T-profile

Except for the lack of support for unreliable messages, time and network management and the differences in TP network and NNN specifications, the services implemented by this T-profile shall be the same as described in 7.

The T-profile shall implement the MAU-LNA link as a reliable message service with one priority level (normal priority) and support one or more stream links (for MAU-MAU stream connections). The following clauses define exceptions from the requirements of 7.

8.3 Reliable stream service

All provisions of 7.4 applies, with the following exceptions:

- Only the normal priority shall be supported (no use of out of band pointer). All requests for CP with other priorities shall be mapped to a normal priority CP.

8.4 Reliable message service

All provisions of 7.5 applies, with the following exceptions:

- Only the normal priority shall be supported. All messages transmitted on the link shall be sent on a normal priority stream link.

8.5 Other services

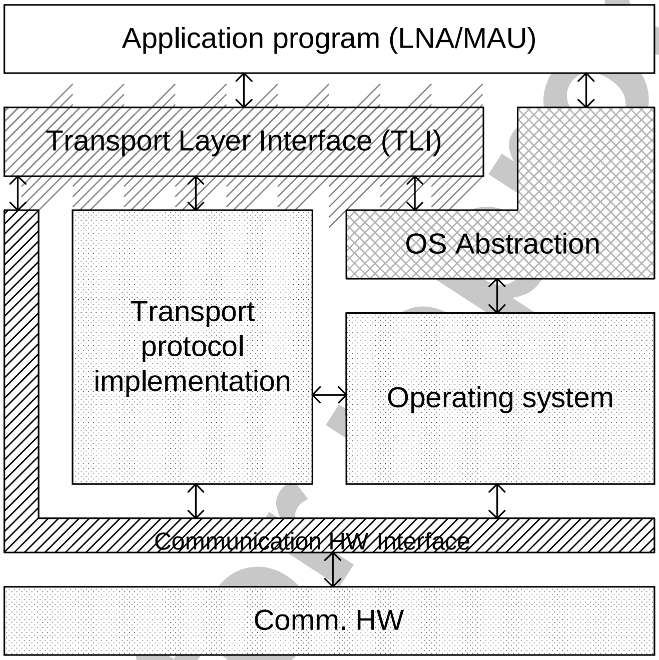
No other services are required from this T-profile.

Annex A
(Informative)

Typical software library structure

The transport protocol interface will be dependent on the operating system in use, while the TLI should be standardised (at least within one software system) to allow the same application software to use different T-profiles.

The software structure for a typical host computer is illustrated in the below figure.


Application program (LNA/MAU)
Transport Layer Interface (TLI)
OS Abstraction
Transport protocol implementation
Operating system
Communication HW Interface
Comm. HW

The transport related software typically consists of the TLI and some extra software to handle special interfaces to the communication hardware, e.g., for priority and redundancy support. The bulk of functionality will be in the existing operating system and transport protocol implementation.

One will also normally see an operating system abstraction module that supplies operating system independent services to the application. This principle is used to enhance portability of application program code. The TLI will in many cases use the same OS abstraction module.

The TLI must be the same for any number of T-profile networks and different T-profiles. The TLI will typically have to be implemented as one generic module (generic in the sense of being special for a certain operating system and software architecture, but generic over T-profiles and TP networks) and one or more T-profile or TP network dependent modules. These modules are strictly speaking not part of the TLI, but need to be tightly integrated with it.

The special variant of the TLI called the LNAC (LNA Communication module) shall supply connectivity between LNA and MAU based on IPC or other communication mechanisms. It is strongly suggested that also the LNAC is created as an interoperable module, where different communication modes can be selected, preferably also during run-time (if, e.g., a MAU shall be submitted to WAN test integration).

Annex B
(Informative)

Channel synchronisation example

The following algorithm is suggested as a possible implementation of a synchronisation mechanism:

The T-profile maintains a descriptor block for each channel with the following information: The current octet count (counted from initial opening of channel) and eight checksums, each with the octet count at which they were made and a flag saying if the checksum has been compared with the other channel.

The checksum is calculated by adding the octets together in a 32 bit word. For each 64 octets, the checksum is shifted one bit to the right (divided by two). A new checksum is saved for each 4096 octets, except when restoring a link failure, where a checksum is stored immediately after the second link was restored (entering RC_SWAIT) and 128 octets thereafter. The latter provision is for a rapid restoration of full redundancy. The checksum is unsigned and an overrun in the checksum is handled as a simple binary addition wrap-around.

When a checksum for one channel is checked against the other, the compared flag is set. This means that the checksum field can be reused. If no reusable fields exist (due to a long delay on one channel), the effective interval (4096) can be doubled. However, the checksum closest to the other channels current octet count shall always be kept. The interval shall be reset as soon as the other channel resumes activity.

The checksum algorithm provides a simple infinite impulse response and should be able to track history fairly well. As messages generally can be expected to be more than 32 octets long, it should be able to trap most errors as specified in the previous clause.

Annex C
(Informative)

NTP and SNTP overview

The Network Time Protocol (NTP) [RFC 1305] defines an architecture for a time service and a protocol to distribute time information. It is used to synchronise the time of a computer client or server to another server or time reference.

NTP is specially designed to achieve high reliability and accuracy even on an Internet which includes paths involving multiple gateways and unreliable networks, and it is included as a standard protocol in the IP suite. NTP is built on the Internet protocol and the user datagram protocol.

NTP time distribution is organised in a hierarchy of NTP servers. At the top of the hierarchy a primary time server is connected to a external source of UTC. NTP time servers can operate in different modes. The mode of each server indicates the behaviour the other machines can expect from it. The broadcast mode is specially intended for use in high speed LANs where the highest accuracy is not required. In broadcast mode, one or more time servers on the LAN periodically broadcasts the time. The interval between the messages can be specified by the `minpoll` subcommand. The workstations then determine the time on the basis of a pre-configured latency specification in the order of a few milliseconds. In order to reliably set the host clock, the client polls the server in burst mode after it has received a message for the first time. At the next broadcast message the client will compute the difference between the system clock and the server time given by the multicast message. This is used to correct for latency in later broadcast messages.

NTP also supports a symmetric peer mode where two or more time servers established on the same hierarchical level (stratum) can synchronise themselves. This synchronisation will handle differences between the servers’ individual time sources and can be used to create a back-up functionality between the servers.

The Simple Network Time Protocol (SNTP) [RFC 2030] is a simplification of the Network Time Protocol. When ultimate performance of the full NTP implementation is not needed or cannot be justified, SNTP can be used. In LAN installations where network latencies are low SNMP may be a good alternative for most clients. This makes for a simpler configuration of clients at the cost of somewhat lower time accuracy. One of the features not found in the SNTP is the support for interfacing to time sources. This means that SNTP can not easily be used to implement a time server connected to a GPS receiver.

A SNTP client can operate in multicast mode, unicast mode or anycast mode. In multicast mode, the client sends no request and waits for a broadcast from a designated multicast server. In unicast mode, the client sends a request to a designated unicast server and expects a reply from that server. In anycast mode, the client sends a request to a designated local broadcast or multicast group address and expects a reply from one or more anycast servers. The client uses the first reply received to establish the particular server for subsequent unicast operations. Unicast mode gives better time accuracy as the client and server are able to estimate transmission delays. With a local area network and good configuration files, the SNMP services may deliver time accuracy to within a few microseconds.

Annex D
(Informative)

Overview of Internet Network Management Services

The Simple Network Management Protocol (SNMP) belongs to the group of Internet application protocols. SNMP is primarily concerned with transfer of control data blocks called MIB (Management Information Base). The purpose of this control flow is to facilitate the necessary flow of management information, status- and statistical data between certain network components and a management system. The scope of this overview is limited to SNMP v1 and MIB v1.

The tasks inherent in network management cannot be reduced to deal with unexpected operational interference alone, but must also comprise observations of significant measuring points within the network to identify imminent symptoms which are likely to indicate emerging problems in network performance. To provide a concise summary with respect to SNMP-related transport services for control data, the following tasks can be identified:

- **Configuration Management:**
  1) Recognition of certain network components.
  2) On-line discovery of complete network topology .
- **Performance Management:**

1) Recognition of significant emerging problems.
2) Alert- and alarm management.
3) Network analysis and –monitoring.
4) Error-processing.

- **Change Management:**

1) Distribution of software (Installation and configuration).
2) Modification of network topology (Move management).
3) Execution of planning tasks.

- **Operations Management:**

1) Execution of classical management operations.
2) Trouble ticketing (Generation of a problem database).

- **Business Management:**

1) Grant of adequate safety of access.
2) Data security.

Within all tasks listed above SNMP may be engaged as transport protocol for control data with respect to network management.

For use in an integrated control system, some of these tasks are of no or of minor interest. For this standard, the most interesting aspects are configuration and performance management.

SNMP Architecture

The SNMP protocol cannot be integrated smoothly into the ISO-OSI layer model since SNMP is connected to every single layer within the protocol stack. SNMP is capable to access management data from every layer in the stack. In the context of this standard, SNMP will only be used for the layers up to transport. Higher layers will be catered to through A-profile management services.

An SNMP management system can be decomposed into three components:

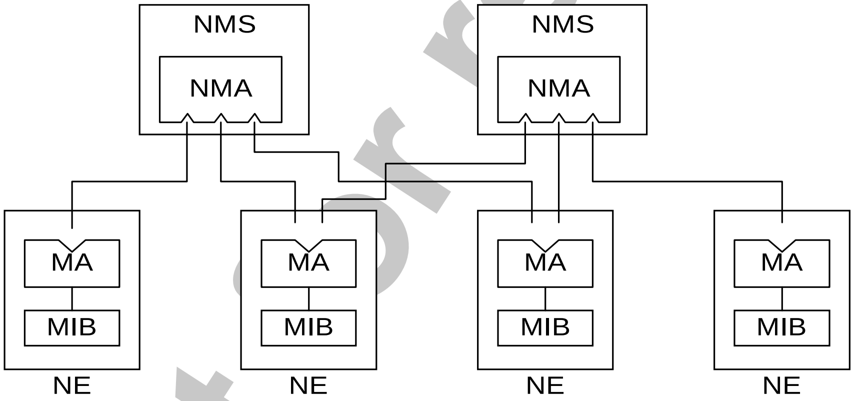
- Network Management Station (NMS)
- SNMP Agents
- Proxy Agents

The NMS is used to monitor the network elements (NE, typically host computers with their protocol entities, routers, switches and gateways). In each of the NE a management agent (MA) is located. The SNMP agents are connected to the NMS via the SNMP protocol.

To transport and represent management-relevant data, three object-categories are used:

- **SMI** (Structure and Identification of Management Information) is a model for description of the structure of needed information..
- **MIB** (Management Information Base) is an agreed standard for description of individual network elements and their sub-objects.
- **SNMP** (Simple Network Management Protocol) is the transport protocol for realisation of communication between **NMS** and the **Management Agents** (MA).

Some of the components are illustrated in the below figure.



Management Information Base (MIB)

SNMP defines a protocol that permits operations on a collection of variables. The following table lists some of the defined variable groups and the number of variables in each group.

Group	Description	No of var.
system	Basic System Information	7
interfaces	Network attachments	23
at	Address Translation	3
ip	Internet Protocol	38
icmp	ICMP Statistics	26
egp	Exterior Gateway Protocol	18
tcp	TCP statistics	19
udp	UDP statistics	7

The MIB defines the objects which may be managed for each layer in the TCP/IP protocol. Each managed node supports only those groups that are appropriate. If a group is appropriate, all objects in that group must be supported. The list of managed objects defined has been derived from those elements considered essential. This approach of taking only the essential objects is not restrictive, since the SMI provides extensibility mechanisms like definition of a new version of the MIB and definition of private or non-standard objects.

Below are some examples of objects in each group; the complete list is defined in RFC 1213.

System Group

- sysDescr - Full description of the system (version, HW, OS)
- sysObjectID - Vendor's object identification
- sysUpTime - Time since last re-initialisation

Interfaces Group

- ifType - Interface type
- ifAdminisStatus - Status of the interface
- ifLastChange - Time the interface entered in the current status
- ifINErrors - Number of inbound packets that contained errors
- ifOutDiscards - Number of outbound packets discarded

Address Translation Group

- atTable - Table of address translation

IP Group

- ipInHdrErrors - Number of input datagrams discarded due to errors in their IP headers
- ipInAddrErrors - Datagrams discarded due to errors in their IP address
- ipInUnknownProtos - Datagrams discarded due to unknown protocol.
- ipReasmOKs - Number of IP datagrams successfully re-assembled
- ipRouteMask – Sub-net mask for route

ICMP Group

- icmpInMsgs - Number of ICMP messages received
- icmpOutErrors - Number of ICMP messages not sent due to problems within ICMP

TCP Group

- tcpMaxConn - Limit on the number of TCP connections the entity can support
- tcpInErrs - Number of segments discarded due to format error
- tcpOutRsts - Number of resets generated

UDP Group

- udpInDatagrams - Number of UDP datagrams delivered to UDP users
- udpOutDatagrams - Number of UDP datagrams sent from this entity

EGP Group

- egpNeighAddr - The IP address of this entry's EGP neighbour
- egpNeighState - The EGP state of the local system with respect to neighbour

This is not the complete MIB definition but it is presented as an example of the objects defined in each group.

Implementation details

SNMP defines a protocol that permits operations on a collection of variables. This set of variables (MIB) and a core set of variables is predefined. The design of the MIB makes provision for extension of this core set. Unfortunately, conventional SNMP agent implementations provide no means for an end user to make new variables available. The SNMP DPI (Distributed Programming Interface) addresses this issue by providing a light-weight mechanism that permits end users to dynamically add, delete, or replace management variables in the local MIB without requiring recompilation of the SNMP agent. This is achieved by writing the so-called subagent that communicates with the agent via the SNMP DPI. It is described by G. Carpenter and B. Wijnen (T.J. Watson Research Center, IBM Corp.) in RFC 1228.

The SNMP DPI allows a process to register the existence of a MIB variable with the SNMP agent. When requests for the variable are received by the SNMP agent, it will pass the query on to the process acting as a subagent. This subagent then returns an appropriate answer to the SNMP agent. The SNMP agent eventually packages an SNMP response packet and sends the answer back to the remote network management station that initiated the request. None of the remote network management stations have any knowledge that the SNMP agent calls on other processes to obtain an answer.

Communication between the SNMP agent and its clients (subagents) takes place over a stream connection. This is typically a TCP connection, but other stream-oriented transport mechanisms can be used (as an example, the VM SNMP agent allows DPI connections over IUCV).

The SNMP Agent DPI can:

- Create and delete sub-trees in the MIB
- Create a register request packet for the subagent to inform the SNMP agent
- Create response packet for the subagent to answer the SNMP agent's request
- Create a TRAP request packet.

____